Lehrstuhl für Kommunikationsnetze
Technische Universität München

IP-Netze mit Interdomain-BGP-Routing: Konvergenzverhalten, Dienstqualität und Dimensionierung

Thomas Schwabe

Vollständiger Abdruck der von der Fakultät für
Elektrotechnik und Informationstechnik der Technischen Universität München
zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr.-Ing. Ulf Schlichtmann
Prüfer der Dissertation	1.	Univ.-Prof. Dr.-Ing. Jörg Eberspächer
	2.	Univ.-Prof. Dr.-Ing. Erwin P. Rathgeb, Universität Duisburg-Essen

Die Dissertation wurde am 13. September 2006 bei der Technischen Universität München
eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik
am 20. März 2007 angenommen.

Vorwort

Diese Arbeit ist während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Kommunikationsnetze der Technischen Universität München entstanden.

Mein besonderer Dank gebührt meinem Doktorvater Herrn Prof. Dr.-Ing. Jörg Eberspächer für das große Engagement, mit dem er meine wissenschaftliche Arbeit begleitet und unterstützt hat. Auch für die angenehme Arbeitsatmosphäre in seiner LKN-Familie, für die großzügige Förderung des fachlichen Austausches mit der nationalen und internationalen Forschungsgemeinschaft sowie für die Möglichkeit, an Industrieprojekten und der Studentenausbildung mitzuarbeiten, möchte ich mich herzlich bedanken.

Herrn Prof. Dr.-Ing. Erwin P. Rathgeb danke ich für sein durch die Übernahme des Zweitgutachtens bekundetes Interesse.

Desweiteren möchte ich meinen Kollegen am Lehrstuhl für die offene und kreative Zusammenarbeit meinen Dank aussprechen, die zum Entstehen dieser Arbeit einen wertvollen Beitrag geleistet hat.

Meinem ehemaligen Kollegen und Projektpartner Claus Gruber gilt mein besonderer Dank. Durch unsere ständigen Diskussionen konnten wir unser Projekt erfolgreich meistern und uns gegenseitig immer wieder neu für die anstehenden Aufgaben motivieren. Danken möchte ich auch Dominic Schupke, Thomas Fischer, Matthias Scheffel, Robert Prinz und Moritz Kiese, die als Mitglieder der Forschungsgruppe “Photonic Networks und Network Resilience Group” mit dem steten fachlichen Austausch mitgeholfen haben, die zahlreichen Probleme beim Fortgang dieser Arbeit und der Projektbearbeitung zu lösen.

Den administrativen Mitarbeitern, insbesondere Herrn Dr. Maier, danke ich für die vielfältige Unterstützung und angenehme Zusammenarbeit.

Eine sehr wichtige Stütze war meine Frau Ulrike, bei der ich mich ganz besonders für ihre Motivation sowie das entgegengebrachte Verständnis bedanke. Meinen Eltern, Maria und Walter Schwabe, danke ich herzlich für ihre vielseitige Unterstützung und Geduld.

München, im September 2006 Thomas Schwabe

Kurzfassung

Der derzeitige Telekommunikationsmarkt ist durch die Konvergenz der Netze gekennzeichnet. Darunter wird das Konzept verstanden, die Daten aller Dienste über einen Netztyp zu übertragen. Für diesen Zweck ist es erforderlich, daß alle beteiligten Netze die verschiedenen Anforderungen an die Dienstqualität gewährleisten können. Als Kerntechnologie wurden die Netze, die das *Internet Protocol* (IP) als Adressierungsprotokoll benutzen, ausgewählt. Es besteht die Herausforderung, die Dienstqualität (*Quality of Service* – QoS) in einem IP-Netz (Intradomain) und zwischen den IP-Netzen (Interdomain) sicherzustellen.

Für die Garantie der Dienstqualität innerhalb eines IP-Netzes existiert eine Vielzahl von Lösungsansätzen. Im Gegensatz dazu gibt es für das Sicherstellen der Dienstqualität zwischen den IP-Netzen kein Konzept, das dies uneingeschränkt leistet. Aus diesem Grund wird in dieser Arbeit ein Konzept entwickelt, das die Interdomain-Dienstqualität garantieren kann und das die besonderen Anforderungen des heutigen Internets, welches die Gesamtheit aller IP-Netze repräsentiert, berücksichtigt.

Gleichbedeutend mit der Zusage der Dienstqualität ist die Sicherstellung einer hohen Fehlerrobustheit der eingesetzten Mechanismen. Deswegen wird im ersten Teil der Arbeit das Konvergenzverhalten des Interdomain-Routing-Protokolls – *Border Gateway Protocol* (BGP) – untersucht und dabei festgestellt, daß im allgemeinen nur die Beschleunigung des Verteilens der Routing-Informationen zu einer kurzen Konvergenzzeit von BGP führt. Für diesen Zweck wird BGP um das *Enhanced Path Vector Protocol* erweitert und gezeigt, daß dadurch das BGP-Zeitglied *Minimum Route Advertisment Interval* deaktiviert werden kann und dies zur Minimierung der Dauer der Konvergenz führt. Diese beiden Mechanismen werden zu dem erweiterten Interdomain-Routing-Protokoll BGP+ zusammengefaßt. Desweiteren werden die notwendigen Anpassungen spezifiziert, um den gleichzeitigen Einsatz von BGP und BGP+ in verschiedenen autonomen Systemen sicherzustellen.

Basierend auf BGP+ wird im zweiten Schritt das Konzept für Garantierte Interdomain-Dienstqualität entwickelt. Dabei wird BGP+ um die Unterscheidung, ob ein Interdomain-Pfad die Dienstqualität gewährleisten kann oder nicht, erweitert und eine Schnittstelle zum Interdomain-Signalisierungsprotokoll *QoS NSIS Signaling Layer Protocol* eingeführt. Besondere Beachtung findet die Sicherstellung der Dienstqualität des Interdomain-QoS-Verkehrs auch nach einer Routing-Änderung. Es werden Dimensionierungsregeln für die Interdomain-Verbindungen aufgestellt, denen eine Abschätzung der Interdomain-Verkehrsverschiebungen zugrunde liegt.

Der letzte Teil der Arbeit untersucht den Einfluß des Interdomain-QoS-Verkehrs auf die Dimensionierung eines Netzes. Die Änderung zu den bestehenden Netzdimensionierungsverfahren besteht in der Herausforderung, daß ein Netzbetreiber garantieren muß, daß der Interdomain-QoS-Verkehr das Netz ohne Einbuße der Dienstqualität und ohne Kenntnis des Ausgangs-Routers passieren kann. Dafür wird ein neues Netzdimensionierungsverfahren vorgeschlagen, das unter Berücksichtigung des Interdomain-Routing bei der Bestimmung der Verkehrsmatrix den Kapazitätsbedarf für die Weiterleitung des Interdomain-QoS-Verkehrs deutlich reduziert.

Inhaltsverzeichnis

Inhaltsverzeichnis **vii**

1 Einführung **1**

1.1 Motivation der Arbeit 1

1.2 Einordnung der Arbeit 2

1.3 Beitrag der Arbeit 3

1.4 Struktur der Arbeit 4

2 Grundlagen und Stand der Technik **5**

2.1 Routing in IP-Netzen 5

2.1.1 Intradomain Routing 6

2.1.2 Interdomain-Routing mit BGP 7

2.1.2.1 Austauschen der BGP-Routing-Informationen 7

2.1.2.2 Bestimmung des besten Pfads 10

2.1.2.3 Beeinflussung der Bestimmung der besten Pfade 11

2.1.2.4 Einteilung der autonomen Systeme und deren Konsequenzen . 12

2.1.2.5 Routen-Reflektoren 14

2.2 Garantie der Dienstqualität in IP-Netzen 16

2.2.1 Parameter zur Beschreibung der qualitativen Eigenschaften eines Dienstes 17

2.2.2 Architekturen für die Garantie der Intradomain Dienstqualität 18

2.2.2.1 Integrated-Services-Architektur 19

2.2.2.2 Differentiated-Services-Architektur 19

2.2.3 Stand der Technik und Ausblick 20

2.3 Netzplanung 21

2.4 Internet-Topologien 24
2.4.1 Eigenschaften der Internet-Topologie 25
2.4.2 Generierung von Internet-Topologien 26
2.4.3 Verwendete Internet-Topologien 27
2.5 Zusammenfassung 29

3 Konvergenzverhalten von BGP 31
3.1 Interdomain-Routing im Internet 31
3.2 Zeitverhalten von BGP bei Topologieänderungen 34
3.2.1 Dauer der BGP-Reaktion auf Topologieänderungen 35
3.2.2 Ablauf der BGP-Konvergenz 36
3.2.2.1 Nichterreichbarkeit eines Präfixes 36
3.2.2.2 Bekanntgabe eines unbekannten Präfixes 37
3.2.2.3 Änderung eines bekannten Pfads 38
3.3 Verbesserung des BGP-Konvergenzverhaltens 38
3.3.1 Schnelle Fehlererkennung 39
3.3.2 Unterdrückung der Oszillation der Pfade 39
3.3.3 Einführung von Redundanz 41
3.3.4 Lokale Ersatzpfade 41
3.3.5 Vermeiden des BGP-Pfadtestens 42
3.3.6 Overlay-Netze 44
3.3.7 Bewertung der existierenden Lösungsvorschläge 44
3.4 Untere Grenze der BGP-Konvergenzzeit 47
3.4.1 Durchlaufzeit 48
3.4.2 Berücksichtigung von Routen-Reflektoren 49
3.4.3 Abschätzung der Durchlaufzeit 50
3.4.4 Abschätzung der BGP-Konvergenzzeit 51
3.4.5 Minimale BGP-Konvergenzzeit 51
3.5 Beschleunigung der BGP-Konvergenz 53
3.5.1 Einfluß des Zeitglieds MRAI 53
3.5.2 Simulationsmodell 54

3.5.2.1 Simulator SSFnet 54
3.5.2.2 Verwendete Simulationsszenarien 54
3.5.2.3 Parameterauswahl 56
3.5.3 Optimale Einstellung des MRAI 56
3.5.3.1 Einfluß der Netzgröße 56
3.5.3.2 Konvergenzverhalten beim Umschalten auf einen Ersatzpfad . 58
3.5.3.3 Untersuchung verschiedener Verarbeitungszeiten 63
3.5.3.4 Einfluß der internen Struktur eines autonomen Systems auf das Konvergenzverhalten 65
3.5.4 Bewertung der erzielten Ergebnisse 68
3.5.5 Einsatz von EPIC 69
3.5.5.1 Fehlerreaktion von EPIC 70
3.5.5.2 Einordnung der Ergebnisse 71
3.6 BGP-Routing mit minimaler Konvergenzzeit 74
3.7 Zusammenfassung 76

4 Konzept für Garantierte Interdomain-Dienstqualität 77
4.1 Rahmenbedingungen 77
4.1.1 QoS-Inseln 78
4.1.2 Routing innerhalb der QoS-Inseln 79
4.2 Bestehende Lösungsvorschläge 80
4.2.1 Das MESCAL-Konzept 80
4.2.2 QoS Enhanced BGP 82
4.2.3 Auf q-BGP aufbauende Ansätze 83
4.2.3.1 Overlay-Netze 83
4.2.3.2 ENTHRONE-Ansatz 84
4.2.4 Interdomain-Ressourcenreservierung mit BGRP und SICAP 84
4.2.5 Protokollfamilie der NSIS-Arbeitsgruppe 85
4.2.6 Bewertung der existierenden Ansätze 86
4.3 Entwurf des KGID 89
4.3.1 Gewährleistung der Interdomain-Dienstqualität 89

4.3.2 Einsatz von BGP+ in den QoS-Inseln 90
4.3.3 Verhalten bei Topologieänderungen 92
4.3.3.1 Kenntnisse über verfügbare Ressourcen innerhalb der QoS-Inseln 92
4.3.3.2 Konsequenz für die Ressourcenreservierungen 93
4.3.3.3 Schnittstelle zum NSIS Transport Layer Protocol 94
4.4 Dimensionierung der Interdomain-Verbindungen 95
4.4.1 Modellierung 96
4.4.2 Ergebnisse 98
4.4.2.1 Stub-NGN 99
4.4.2.2 Kleine Tier2-NGN 100
4.4.2.3 Große Tier2-NGN 104
4.4.2.4 Transit-NGN 107
4.4.2.5 Dimensionierungsregeln für das KGID 110
4.4.3 Auswirkungen auf die Netzzugangskontrollverfahren 112
4.5 Zusammenfassung 113

5 Intradomain-Netzdimensionierung für QoS-Verkehr 115
5.1 Transport von Interdomain-QoS-Verkehr 115
5.1.1 Nicht robuste Netzdimensionierung 116
5.1.2 Robuste Netzdimensionierung 116
5.1.3 Netzdimensionierung unter Verwendung von Netzzugangskontrollverfahren 117
5.1.4 Statistische Netzdimensionierung 118
5.2 Dimensionierung mit Rücksicht auf das BGP+-Routing 119
5.2.1 Einteilung des Verkehrs 120
5.2.2 Ermitteln der Verkehrseinteilung 120
5.2.3 Einschränkung der möglichen Ausgangs-Router 121
5.3 Dimensionierung eines NGN 124
5.3.1 Untersuchte Szenarien 124
5.3.2 Verwendete Verkehrsmatrizen 124
5.3.3 Eingesetztes Routing-Verfahren 126

5.3.3.1 Verwendung eines genetischen Algorithmus als Heuristik . . . 127
5.3.3.2 Eingesetzter genetischer Algorithmus 128
5.3.4 Ergebnisse der BGP+-Dimensionierung 129
5.3.5 Eingeschränkt-robuste BGP+-Dimensionierung 131
5.3.5.1 Eingeschränkt-robuste BGP+-Dimensionierung für ein Transit-NGN 132
5.3.5.2 Eingeschränkt-robuste BGP+-Dimensionierung für ein Tier2-NGN 133
5.3.6 Anwendung der Dimensionierungsregeln des KGID 134
5.3.7 Änderung bei der Netzplanung eines NGN 136
5.4 Zusammenfassung 137

6 Zusammenfassung **139**

Abbildungsverzeichnis **143**

Tabellenverzeichnis **147**

Literaturverzeichnis **149**

Abkürzungsverzeichnis

BGP Border Gateway Protocol

CAC Call Admission Control

CAPEX Capital Expenditure

CIDR Classless Interdomain Routing

DiffServ Differentiated Services

DLZ Durchlaufzeit*

DSCP DiffServ Code Point

E-BGP Exterior BGP

EPIC Enhanced Path Vector Protocol

FIB Forwarding Information Base

GIST General Internet Signaling Transport

GLP Generalized Linear Preference

I-BGP Interior BGP

IETF Internet Engineering Task Force

IntServ Integrated Services

IP Internet Protocol

KE Kapazitätseinheiten*

KGID Konzept für Garantierte Interdomain Dienstqualität*

Loc-RIB Local RIB

MED Multiple Exit Discriminator

MESCAL Management of End-to-end Quality of Service Across the Internet at Large

MPLS Multi Protocol Label Switching

MRAI Minimum Route Advertisment Interval

MRRI Minimum Resource Reservation Interval*

MTTF Mean Time To Fail

MTTR Mean Time To Repair

NAC Network Admission Control

NGN Next Generation Network

NSIS Next Steps In Signaling

NSLP NSIS Signaling Layer Protocol

NTLP NSIS Transport Layer Protocol

OSPF Open Shortest Path First

OPEX Operational Expenditure

q-BGP QoS enhanced BGP

QoS Quality of Service

RFD Route Flap Damping

RIB Routing Information Base

RR Routen-Reflektor*

RSVP Resource Reservation Protocol

SLA Service Level Agreement

* Eigene Abkürzungen

Kapitel 1

Einführung

1.1 Motivation der Arbeit

Mit der Deregulierung der Telekommunikationsmärkte wurde ein Wettbewerb zwischen bestehenden und neu entstehenden Netzbetreibern initiiert. Von Beginn an resultierte aus diesem Wettbewerb fallende Preise bei den Telefondiensten. Mittlerweile hat sich der Wettbewerb auf die breitbandigen Internetzugänge ausgeweitet. Die Suche nach neuen Geschäftsmöglichkeiten und die stetige Erhöhung der Bandbreite der Internetzugänge hat dazu geführt, daß diese Breitbandzugänge für das Angebot neuer Dienste genutzt werden. Neben dem reinen Internetzugang kann nun ein Internet Service Provider auch Telefon- oder Fernsehdienste zur Verfügung stellen, genauso wie ein Kabelnetzbetreiber neben Fernsehdiensten auch Sprach- und Datendienste anbieten kann.

Das erweiterte Angebot mehrerer Dienste aus der Hand eines Netzbetreibers und der damit einhergehende wachsende Wettbewerb haben den Kostendruck für die Bereitstellung dieser Dienste noch weiter erhöht. Um diesem zu begegnen, versuchen die Netzbetreiber, alle Dienste über ein Netz bereitzustellen, und verabschieden sich damit von der Philosophie, für jeden Dienst ein genau auf dessen Bedürfnisse abgestimmtes Netz zu errichten [Ebe06]. Dieser Prozeß wird unter dem Begriff "Konvergenz der Netze" zusammengefaßt.

Als Technologieträger, über die die Datenübertragung aller Dienste erfolgen kann, haben sich die Netze, die auf das *Internet Protocol* (IP) aufbauen, durchgesetzt. Allerdings werden neue Mechanismen für diese IP-Netze benötigt, um den Anforderungen aller Dienste gerecht zu werden. Daneben muß auch weiterhin die Datenübertragung der klassischen Dienste, wie das *World Wide Web* (WWW) und die Email-Kommunikation, unterstützt werden. Außerdem muß sichergestellt werden, daß das Internet weiterhin offen für neue Dienste bleibt, z.B. für die in den letzten Jahren extrem populär gewordenen *Peer-to-Peer* Anwendungen.

Als Konzept für die Unterstützung der bestehenden und zukünftigen Dienste hat sich die Klassifizierung anhand der Anforderungen der spezifischen Anwendungen an die Dienstqualität (*Quality of Service* – QoS) durchgesetzt. Dabei fordert jede Anwendung die benötigte Dienstqualität an,

und die IP-Netze sollten in der Lage sein, die geforderte Dienstqualität sicherzustellen. Dafür sind in den letzten Jahren viele verschiedene Ansätze entwickelt worden. Allerdings konzentrieren sich die meisten auf die Gewährleistung der Dienstqualität innerhalb eines IP-Netzes. Deswegen besteht die Herausforderung darin, auch über Netzgrenzen hinweg die Garantie der Dienstqualität zu ermöglichen. Zur Lösung dieser Aufgabe leistet diese Arbeit ihren Beitrag.

1.2 Einordnung der Arbeit

Generell kann die Kommunikation in IP-Netzen in zwei verschiedene Sichtweisen unterteilt werden. Betrachtet man nur ein einzelnes IP-Netz, dann wird dies als Intradomain-Kommunikation bezeichnet. Im Gegensatz dazu versteht man unter der Interdomain-Kommunikation die Kommunikation zwischen verschiedenen IP-Netzen. Zur Vereinfachung werden meistens die einzelnen Netze oder *Autonomen Systeme* (AS) zu einem Knoten zusammengefaßt, wie sie als blaue Wolken in Abbildung 1.1 veranschaulicht sind.

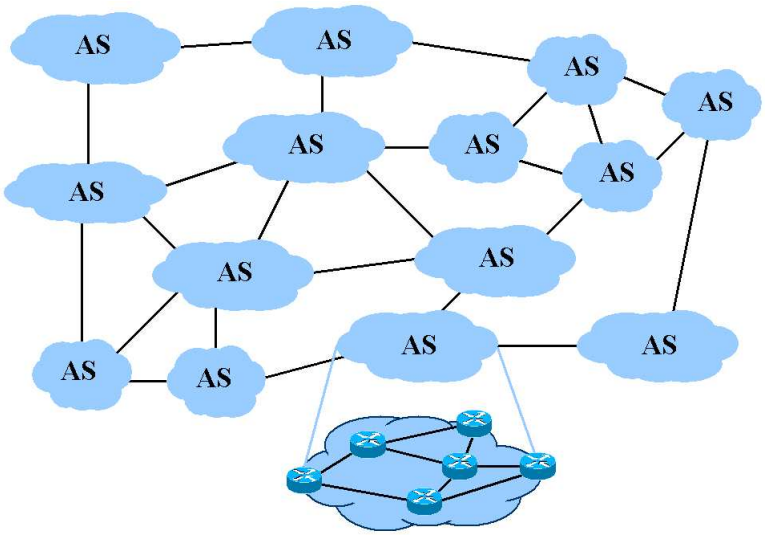


Abbildung 1.1: Intra- und Interdomain-Kommunikation

Wird der Fokus auf die Intradomain-Kommunikation gelegt, wird ein autonomes System zu einem Netz mit mehreren untereinander verbundenen Knoten erweitert und die Untersuchung auf diesen Teil beschränkt. Dies wird beispielhaft durch die detaillierte Darstellung des unteren autonomen Systems in Abbildung 1.1 demonstriert.

Der Schwerpunkt dieser Arbeit liegt auf der Interdomain-Kommunikation zwischen IP-Netzen, denen die Eigenschaften des heutigen Internets zu Grunde gelegt werden. Allerdings ist es notwendig, die Kommunikation innerhalb eines Netzes mit in die Betrachtungen einzubeziehen, um eine durchgehende Ende-zu-Ende-Kommunikation zwischen Teilnehmern, die verschiedenen Netzen zuzuordnen sind, zu ermöglichen.

Ein Hauptgrund für die getrennte Sichtweise auf die Kommunikation innerhalb eines Netzes und zwischen mehreren Netzen liegt unter anderem an den unterschiedlichen Eigenschaften der Netze und den sich daraus ergebenden Freiheitsgraden für die Gestaltung und Konfiguration der Intra- bzw. Interdomain-Kommunikation.

Untersucht man nur ein einzelnes Netz, hat dies den Vorteil, daß dieses Netz von einer administrativen Einheit betrieben wird und sich somit das Wissen über das Netz in einer Hand befindet. Das kann beispielsweise benutzt werden, um das Verhalten des Netzes für festgelegte Fehlerbilder im voraus zu berechnen oder das Netz exakt an veränderte Bedingungen, wie z.B. schwankenden Verkehr, anzupassen.

Auf die Interdomain-Kommunikation, wie sie in dieser Arbeit betrachtet wird, trifft genau das Gegenteil zu. Hier tauschen Netze Daten aus, die von verschiedenen Betreibern unterhalten werden. Zum Teil stehen diese sogar im direkten Wettbewerb, und das Ziel besteht darin, die Kommunikation so zu gestalten, daß der Betreiber eines Netzes die maximale Freiheit bei der Konfiguration und beim Betrieb seines Netzes erreicht.

Diese Arbeitsweise, welche eines der Grundprinzipien der Interdomain-Kommunikation im heutigen Internet darstellt, ist auch der wesentliche Unterschied zu anderen Arbeiten, wie z.B. [Gla03, Tro05], wo ebenfalls die Kommunikation zwischen mehreren Netzen untersucht wird, aber davon ausgegangen wird, daß man das komplette Wissen aller betrachteten Netze besitzt.

Eine Konsequenz der Betrachtungsweise dieser Arbeit ist die Behandlung der Interdomain-Kommunikation in einem großen verteilten System, wo keine zentrale Instanz mit netzübergreifendem Wissen existiert. Die Freiheit jedes Netzbetreibers und die teilweise Konkurrenz führen dazu, daß nur die absolut notwendigen Informationen zwischen den Netzen ausgetauscht werden und das Verhalten im voraus aufgrund des mangelnden Wissens nur sehr bedingt vorhergesagt werden kann. Unter diesen Voraussetzungen wird in dieser Arbeit ein Konzept entwickelt, das die Garantie der Dienstqualität auch zwischen verschiedenen IP-Netzen ermöglicht.

1.3 Beitrag der Arbeit

Das Ziel dieser Arbeit ist die Bereitstellung eines Konzepts, das die Evolution des heutigen Internets zu einem Netz mit Gewährleistung der Dienstqualität zwischen allen autonomen Systemen ermöglicht. Dabei findet die Skalierbarkeitsanforderung an dieses Konzept aufgrund der Größe und des Wachstums des heutigen Internets besondere Beachtung.

Eine wichtige Voraussetzung, um überhaupt die Dienstqualität sicherstellen zu können, ist ein Routing, das eine Fehlerreaktion mit einer kurzen Konvergenzzeit aufweist. Nur so kann gewährleistet werden, daß Routing-Änderungen nicht zum Abbruch der Datenübertragung wegen mangelnder Sicherstellung der Dienstqualität führen. Innerhalb eines Netzes stellt dies kein Problem dar, aber das heutige Interdomain-Routing mit dem *Border Gateway Protocol* (BGP) benötigt für eine Fehlerreaktion mehrere Minuten, bis alle autonomen Systeme den aufgetretenen Fehler verarbeitet und ein endgültiges Ergebnis gefunden haben.

Deswegen wird im ersten Teil das Konvergenzverhalten von BGP umfassend untersucht und festgestellt, daß nur die Beschleunigung des Verteilens der BGP-Routing-Informationen zu einer wesentlich verkürzten Konvergenzzeit führt. Basierend auf der Simulation des Konvergenzverhaltens wird die mögliche Reduzierung der Konvergenzzeit von BGP ermittelt, und die Qualität der Ergebnisse wird anhand der theoretisch abgeschätzten untere Grenze bewertet. Die erzielten Erkenntnisse, um die Konvergenzzeit zu minimieren, werden in dem erweiterten neuen Interdomain-Routing-Protokoll BGP+ umgesetzt.

Aufbauend auf dem BGP+-Interdomain-Routing wird das Konzept für die Gewährleistung der Interdomain-Dienstqualität entworfen. Ein wichtiger Bestandteil dieses Ansatzes ist die Einführungsstrategie, die die Weiterentwicklung des heutigen Internets hin zu einem Netz mit garantierter Interdomain-Dienstqualität vorsieht. Im Rahmen dessen wird BGP+ um einen Mechanismus zur Erfüllung der Anforderungen der Interdomain-Dienstqualität und um eine Schnittstelle mit dem Protokoll für die Interdomain-Ressourcenreservierungen erweitert.

Das neue Konzept ermöglicht es im Gegensatz zu anderen bestehenden Ansätzen, die Dienstqualität auch nach einer Interdomain-Routing-Änderung weiterhin zu gewährleisten. Für diesen Zweck werden in dieser Arbeit Dimensionierungsregeln für die Interdomain-Verbindungen erarbeitet, und dabei werden auch mögliche Netzzugangskontrollverfahren in den autonomen Systemen berücksichtigt.

Der letzte Teil dieser Arbeit beschäftigt sich mit dem Einfluß des Interdomain-QoS-Verkehrs auf die Netzplanung eines einzelnen autonomen Systems. Um den Anforderungen des Interdomain-QoS-Verkehrs gerecht zu werden, wird ein neues Verfahren zur Netzdimensionierung vorgeschlagen. Das Einbeziehen des Interdomain-Routing erlaubt es, den Kapazitätsbedarf eines Netzes für den Interdomain-QoS-Verkehr im Vergleich zu den bestehenden Verfahren deutlich zu reduzieren.

1.4 Struktur der Arbeit

Die nachfolgenden Untersuchungen sind in fünf Kapitel unterteilt. Im Anschluß wird in Kapitel 2 ein Überblick über die für diese Arbeit wichtigen Mechanismen und Protokolle von IP-Netzen gegeben.

Kapitel 3 beschäftigt sich mit dem Konvergenzverhalten des Interdomain-Routing im heutigen Internet und schlägt das verbesserte Interdomain-Routing-Protokoll BGP+ vor.

In Kapitel 4 wird das Konzept für die Gewährleistung der Dienstqualität für Interdomain-Verkehr entworfen und Dimensionierungsregeln für die Interdomain-Verbindungen erarbeitet, um nach einer Interdomain-Routing-Änderung weiterhin die Dienstqualität sicherzustellen.

Das neue Netzdimensionierungsverfahren, das den Einfluß des Interdomain-QoS-Verkehrs mit berücksichtigt, wird in Kapitel 5 erläutert, und die Ergebnisse dieser Arbeit werden in Kapitel 6 zusammengefaßt.

Kapitel 2

Grundlagen und Stand der Technik

Dieses Kapitel beschreibt die Grundlagen für die in den folgenden Kapiteln untersuchten Problemstellungen. Im ersten Abschnitt werden die verschiedenen Routing-Konzepte in IP-Netzen und die dazu gehörigen Protokolle vorgestellt. Dabei wird auf das Interdomain-Routing-Protokoll BGP im Detail eingegangen. Dieses stellt das Standard-Interdomain-Protokoll für das Routing zwischen den verschiedenen Netzen des Internets dar, und dient damit als Grundlage für die Entwicklung von BGP+ in Kapitel 3.

Der zweite Abschnitt beschäftigt sich mit den grundlegenden Prinzipien für die Gewährleistung der Dienstqualität innerhalb eines IP-Netzes. Einige der dabei vorgestellten Mechanismen werden auch für die Garantie der Interdomain-Dienstqualität weiterverwendet.

Daran schließt sich die Betrachtung des prinzipiellen Ablaufs der Netzplanung und -dimensionierung an, die im Kapitel 5 anhand eines Beispiels ausgeführt wird. Zum Schluß werden die in den folgenden drei Kapiteln verwendeten Internet-Topologien beschrieben.

2.1 Routing in IP-Netzen

Um Daten zwischen mehreren Knoten eines IP-Netzes oder zwischen IP-Netzen auszutauschen, benötigt jeder Knoten Informationen, wie er die entsprechenden Ziele erreichen kann. Für diese Aufgabe sind in IP-Netzen die Routing-Protokolle verantwortlich.

Grundsätzlich kann man die heute eingesetzten Routing-Protokolle in *Link-State*- und *Distance-Vector*-Routing-Protokolle einteilen. Daneben werden Routing-Protokolle auch anhand ihrer Einsatzgebiete unterschieden. Die Intradomain-Routing-Protokolle werden innerhalb eines Netzes benutzt. Im Gegensatz dazu versteht man unter Interdomain-Routing-Protokollen diejenigen, die zwischen verschiedenen Netzen eingesetzt werden.

Beide Typen von Routing-Protokollen kommen nebeneinander im Internet zum Einsatz. Unter dem Begriff Internet wird eine Vielzahl von IP-Netzen verstanden, die alle miteinander verbunden sind und IP [Pos81a] als Protokoll für die Adressierung verwenden. Laut [Tel05] besteht

das Internet derzeit aus mehr als 20.000 einzelnen IP-Netzen oder autonomen Systemen. Unter einem autonomen System wird ein IP-Netz verstanden, welches Bestandteil des Internets ist und von einer administrativen Einheit – dem Netzbetreiber – kontrolliert wird.

2.1.1 Intradomain Routing

Das Routing innerhalb eines autonomen Systems ist durch zwei entscheidende Vorteile – die vollständige Topologiekenntnis und die relativ kleine Anzahl an Netzknoten – gekennzeichnet. Aufgrund dessen können auch Link-State-Routing-Protokolle eingesetzt werden, bei denen jeder Netzknoten die komplette Topologie kennt und dieses Wissen für die Berechnung der Pfade zu einem Ziel benutzt.

Auch Distance-Vector-Routing-Protokolle kommen innerhalb eines Netzes zum Einsatz. Diese nutzen aber den Vorteil der vollständigen Topologiekenntnis nicht, da hierbei nur Vektoren, die die Distanz zu einem Ziel enthalten, zur Pfadbestimmung benutzt und zwischen den Netzknoten ausgetauscht werden.

Die heute am häufigsten benutzten Intradomain-Routing-Protokolle sind die folgenden:

- *Routing Information Protocol* (RIP) [Hed88]
- *Open Shortest Path First* (OSPF) [Moy98]
- *Intermediate System - Intermediate System* (IS-IS) [Ora90]
- *Interior Gateway Routing Protocol* (IGRP) [Zin01]
- *Enhanced Interior Gateway Routing Protocol* (EIGRP) [AGLAB94]
- *Multi Protocol Label Switching* (MPLS) [RVC01]

RIP ist das einzige Distance-Vector-Routing-Protokoll in dieser Auswahl. Allerdings ist es aufgrund seiner einfachen Auslegung nur für Netze mit sehr wenigen Netzknoten geeignet. OSPF, IS-IS, IGRP und EIGRP sind Link-State-Routing-Protokolle, die den kürzesten Pfad zwischen zwei Netzknoten berechnen. OSPF und IS-IS bestimmen anhand einer frei wählbaren Metrik die kürzesten Pfade. Bei beiden kann nur ein Wert für jede Verbindung vergeben werden, und damit können nur anhand einer Metrik die kürzesten Pfade berechnet werden. Ansonsten sind sich OSPF und IS-IS sehr ähnlich. Der Hauptunterschied ist die Organisation, die diese Protokolle standardisiert hat. OSPF wurde von der *Internet Engineering Task Force* (IETF) definiert und IS-IS wurde bei der *International Standardisation Organisation* (ISO) entwickelt. IGRP und EIGRP sind zwei proprietäre Protokolle des Router-Herstellers Cisco. Sie bieten die Möglichkeit mehrere Metriken gleichzeitig zu benutzen. Für jede Verbindung können die vier Parameter Bandbreite, Verzögerung, Auslastung und Zuverlässigkeit als eigenständiges Verbindungsgewicht verwendet werden. Anhand dieser sind IGRP und EIGRP in der Lage, für jeden dieser vier Parameter einen eigenen kürzesten Pfad zu einem Ziel zu berechnen.

Bei den bisher beschriebenen Routing-Protokollen entscheidet ein Netzknoten anhand der Zieladresse, über welchen kürzesten Pfad die entsprechenden Pakete weitergeleitet werden. MPLS hebt die Einschränkungen dieses zielbasierten Routing auf. Damit ist MPLS kein IP-Routing-Protokoll im herkömmlichen Sinne und ist somit auch kein Distance-Vector- bzw. Link-State-Routing-Protokoll.

Über welchen Pfad die IP-Pakete weitergeleitet werden, wird bei MPLS anhand eines *Label* entschieden. Für diesen Zweck müssen die Pfade und die dafür vergebenen Label im voraus berechnet werden. Der Vorteil dieser Funktionsweise ist eine größere Freiheit bei der Pfadauswahl zu einem Zielknoten. Allerdings erhöht sich damit auch die Komplexität von MPLS bei der Bestimmung der Pfade. Ein weiterer Vorteil von MPLS ist die schnelle Paketweiterleitung in einem Knoten, da die Forwarding-Tabellen aufgrund der Label deutlich kleiner sind als die Routing-Tabellen der anderen IP-Routing-Protokolle.

2.1.2 Interdomain-Routing mit BGP

Im Gegensatz zu den im vorherigen Abschnitt beschriebenen Routing-Protokollen ist ein einziges Interdomain-Routing-Protokoll für die Pfadsuche zwischen verschiedenen Netzen zuständig. Nur wenn alle Netze dasselbe Routing-Protokoll einsetzen, ist gewährleistet, daß jedes Netz mit jedem kommunizieren kann. Im heutigen Internet hat sich BGP [RLH06] als das Interdomain-Routing-Protokoll durchgesetzt und bildet somit die Basis für die Interdomain-Kommunikation. Damit hat das Interdomain-Routing auch entscheidenden Einfluß auf die Evolution des Internets hin zu einem Netz, das die Interdomain-Dienstqualität gewährleisten kann. Deswegen wird im folgenden die grundlegende Funktionsweise von BGP beschrieben.

Die Skalierbarkeitsanforderung, die aufgrund der Größe sowie des stetigen Wachstums des Internets entsteht, erfüllt nur ein Distance-Vector-Routing-Protokoll. Ein Link-State-Routing-Protokoll kann dies nicht leisten, da die Verwaltung der Verbindungszustände der Skalierbarkeit Grenzen setzt. Aufgrund dessen hat der Vektor mit der Beschreibung der BGP-Pfade als Ziel keine einzelne IP-Adresse, sondern nur eine Gruppe von IP-Adressen. Diese Gruppe wird auch als Präfix bezeichnet, die bei BGP der Version 4 nicht mehr mit den traditionellen IP-Klassen (A, B, C), sondern mit dem so genannten *Classless Interdomain Routing* (CIDR) [FLY+93] aggregiert werden. Dabei besteht ein Präfix aus einer IP-Adresse und der Länge des Präfixes, die angibt, wie viele der ersten 32 Bits fest vergeben sind. Der variable Teil beschreibt die Gruppe der IP-Adressen, die das Präfix repräsentiert. Beispielsweise beschreibt 129.187.128.0/22 den Adreßbereich von 129.187.128.0 bis 129.187.131.255, wobei bei diesem Präfix die letzten 10 Bit bzw. 1024 IP-Adressen variabel sind.

2.1.2.1 Austauschen der BGP-Routing-Informationen

Um die Routing-Informationen zwischen zwei BGP-Routern auszutauschen, sind in [RLH06] vier Nachrichtentypen – *OPEN*, *KEEPALIVE*, *NOTIFICATION* und *UPDATE* – definiert. Zu

Beginn muß zwischen den benachbarten BGP-Routern eine BGP-Verbindung aufgebaut werden. Das wird durch das Abschicken einer BGP-OPEN-Nachrichten initiiert. Der Erhalt einer BGP-OPEN-Nachrichten wird mit zwei BGP-KEEPALIVE-Nachrichten, jeweils eine in jede Richtung, bestätigt.

Für den Austausch der BGP-Nachrichten wird das *Transmission Control Protocol* (TCP) [Pos81b] als Transportprotokoll benutzt und damit wird eine gesicherte Übertragung der BGP-Nachrichten erreicht.

Als BGP-Verbindung wird in dieser Arbeit die Kommunikation zweier BGP-Router untereinander bezeichnet, wohingegen der Begriff Verbindung verwendet wird, um eine physikalische Leitung zwischen zwei Knoten, über die IP-Pakete transportiert werden können, zu beschreiben.

Nach dem erfolgreichen BGP-Verbindungsaufbau wechselt der BGP-Router in den Zustand "Normalbetrieb" (s. Abbildung 2.1), in dem der Austausch der BGP-Routing-Informationen stattfindet. Dieser Zustand wird durch das periodische Senden von BGP-KEEPALIVE-Nachrichten überwacht. Empfängt ein BGP-Router innerhalb einer gewissen Zeit keine BGP-KEEPALIVE-Nachricht oder keine BGP-UPDATE-Nachricht, geht er davon aus, daß die BGP-Verbindung ausgefallen ist. Diese Funktionsweise entspricht dem Prinzip eines *Hello*-Protokolls.

Tritt ein Fehler innerhalb der BGP-Kommunikation auf, wird dieser mit einer BGP-NOTIFICATION-Nachricht dem jeweiligen Nachbar mitgeteilt. Je nach dem aktuellen Zustand von BGP führt dies zu unterschiedlichen Reaktionen, die alle im BGP-Zustandsdiagramm in Abbildung 2.1 aufgeführt sind.

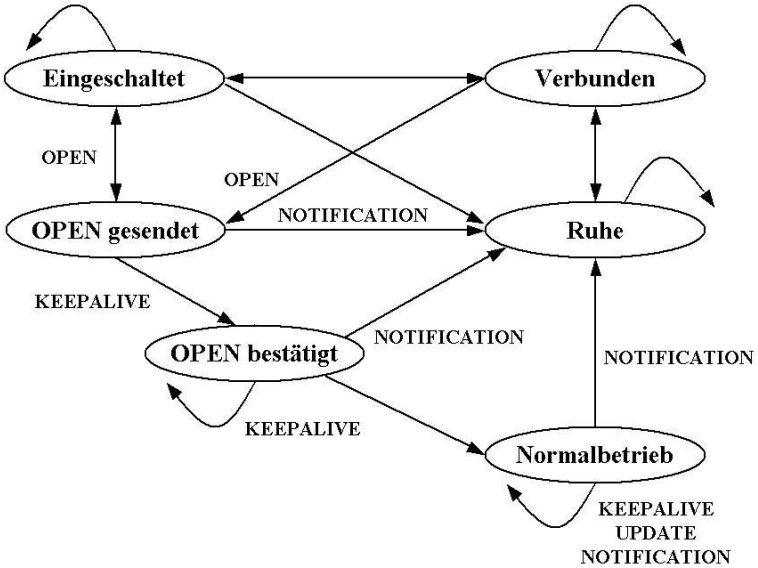


Abbildung 2.1: BGP-Zustandsdiagramm nach [Hal97]

Für den Transport der Routing-Informationen ist die BGP-UPDATE-Nachricht verantwortlich. Normalerweise werden mit einer BGP-UPDATE-Nachricht nur die Änderungen des besten Pfads übermittelt, um damit die Menge der zu übertragenden Routing-Informationen zu reduzieren.

Eine Ausnahme bildet der Aufbau einer BGP-Verbindung. In diesem Fall werden alle bekannten besten Pfade zu allen erreichbaren Zielen an den Nachbar-BGP-Router übermittelt.

Mit der BGP-UPDATE-Nachricht werden zwei verschiedene BGP-Routing-Informationen übertragen. Am Anfang der Nachricht werden die Präfixe, die nicht mehr erreichbar sind, als einfache Liste übermittelt. Diese bezeichnet man als *Withdrawn*. Im zweiten Teil werden die geänderten Pfade als Vektor mitgeteilt, welcher auch *Announcement* genannt werden. Wie diese beiden Teile in eine BGP-UPDATE-Nachricht eingebaut werden, ist in Abbildung 2.2 zu sehen.

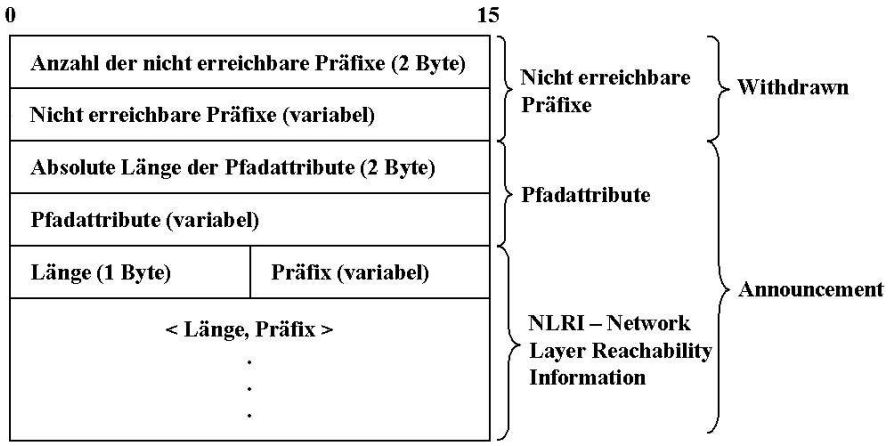


Abbildung 2.2: Aufbau einer BGP-UPDATE-Nachricht nach [Hal97]

Der im Announcement-Teil übertragene Pfadvektor besteht aus dem Ziel des Pfads als Präfix in der CIDR-Notifikation und aus den dem Pfad zugeordneten Attributen. Ein Teil dieser Attribute muß in jedem BGP-Pfadvektor gesetzt sein. Zu den wichtigsten Pflichtattributen zählen der AS-Pfad (*AS_PATH*), die Quelle (*ORIGIN*) und der nächste Knoten des Pfads (*NEXT_HOP*). Zusätzlich sind auch noch optionale Attribute, die gesetzt werden können, aber nicht müssen, in [RLH06] definiert. Häufig verwendete optionale Attribute sind die lokale Präferenz (*LOCAL_PREF*) bzw. der *Multiple Exit Discriminator* (MED). Im Unterschied zu den Pflichtattributen muß nicht jeder BGP-Router alle optionalen Attribute kennen. Erhält ein BGP-Router eine BGP-UPDATE-Nachricht, die einen BGP-Pfadvektor mit einem unbekannten optionalen Attribut enthält, dann verwirft der BGP-Router dieses unbekannte Attribut. Abhängig von der Definition der optionalen Attribute können diese an die Knoten der benachbarten autonomen Systeme weitergegeben, nur innerhalb eines autonomen Systems verwendet oder nur an bestimmte autonome Systeme weitergeleitet werden. Eine detaillierte Beschreibung der von BGP verwendeten Attribute und deren Geltungsbereich ist in [Hal97] zu finden.

Um die Menge der übertragenen BGP-UPDATE-Nachrichten bei der Reaktion auf eine Pfadänderung zu reduzieren, wurde das Zeitglied Minimum Route Advertisment Interval (MRAI) in [RLH06] definiert. Das MRAI sorgt dafür, daß ein BGP-Router eine gewisse Zeit wartet, bevor er die nächste BGP-UPDATE-Nachricht mit demselben Präfix versendet. In Abhängigkeit zwischen welchen BGP-Routern die BGP-UPDATE-Nachrichten ausgetauscht werden, werden in [RLH06] verschiedene Standardwerte vorgeschlagen. Bei der BGP-Kommunikation zwischen

benachbarten autonomen Systemen, die als *Exterior BGP* (E-BGP) bezeichnet wird, beträgt der Standardwert 30 Sekunden. Innerhalb eines autonomen Systems – bei der *Interior BGP* (I-BGP) Kommunikation – ist der Standardwert des MRAI auf 5 Sekunden festgelegt. Auf die Wirkungsweise und auf die mögliche Verbesserung der BGP-Fehlerreaktion durch das MRAI wird in Abschnitt 3.5.1 noch genauer eingegangen.

2.1.2.2 Bestimmung des besten Pfads

Anhand der Attribute der BGP-Pfadvektoren bestimmt der BGP-Routing-Prozeß die besten Pfade zu den entsprechenden Präfixen. Diese Auswahl führt jeder BGP-Router eines autonomen Systems unabhängig von allen anderen BGP-Routern aus. Anhand welcher Priorität ein BGP-Router seinen besten Pfad bestimmt, beschreibt die folgende Aufzählung aus [Hal97]. Dabei repräsentiert die Reihenfolge auch die Priorität der Abfrage.

1. Pfad ignorieren, wenn der als NEXT_HOP beschriebene BGP-Router nicht erreichbar ist.
2. Pfad mit dem höchsten Gewicht (nur bei Cisco Routern) bevorzugen.
3. Pfad mit dem höheren Wert des Attributs LOCAL_PREF bevorzugen.
4. Pfad mit der kleineren Anzahl der autonomen Systeme im Pfadattribut AS_PATH bevorzugen.
5. Pfad anhand des Pfadattributs ORIGIN unterscheiden (1. Quelle liegt innerhalb des autonomen Systems; 2. Quelle liegt außerhalb des autonomen Systems; 3. Quelle ist unbekannt).
6. Pfad mit dem kleineren Wert des Attributs MED bevorzugen.
7. Pfad bevorzugen, der von einem externen gegenüber einem internen BGP-Router empfangen wurde.
8. Pfad mit kürzerem Intradomain-Pfad zu einem BGP-Router bevorzugen.
9. Endgültiges Entscheidungskriterium, z.B. kleinere Identifikationsnummer des Routers, von dem die Nachricht empfangen wurde.

Da die Attribute eines Pfadvektors konfigurierbar sind und bei jedem BGP-Router individuell festgelegt werden können, erhält der Betreiber eines autonomen Systems damit ein wichtiges Werkzeug, um das BGP-Routing an seine Anforderungen anzupassen. Die Regeln, nach denen die Werte der Attribute gesetzt werden, sind ein Bestandteil des Regelwerks, das als *Policies* bezeichnet wird.

2.1.2.3 Beeinflussung der Bestimmung der besten Pfade

Der zweite Bestandteil der Policies sind die Filterregeln, mit denen ein Betreiber individuell bestimmte BGP-UPDATE-Nachrichten herausfiltern kann. Die Kombination aus dem Filtern und Ändern der Attribute der BGP-UPDATE-Nachrichten ergibt ein sehr mächtiges Werkzeug, um die Verarbeitung der BGP-UPDATE-Nachrichten innerhalb seines autonomen Systems zu beeinflussen.

Wie ein Betreiber mittels der Policies in die Bestimmung des besten Pfads Einfluß nehmen kann, wird deutlich, wenn man sich den Ablauf der Verarbeitung einer BGP-UPDATE-Nachricht in einem BGP-Router anschaut. Dafür existieren in einem BGP-Router die folgenden vier BGP-Routing-Tabellen:

- *Adjacent Routing Information Base* am Eingang des Routers – Adj-RIB-IN
- *Local Routing Information Base* – Loc-RIB
- *Forwarding Information Base* – FIB
- *Adjacent Routing Information Base* am Ausgang des Routers – Adj-RIB-OUT

Als erstes werden die via der BGP-UPDATE-Nachrichten neu eingetroffenen Routing-Informationen in den Adj-RIB-IN abgelegt. In jedem BGP-Router existiert jeweils eine Adj-RIB-IN für jede BGP-Verbindung dieses Routers. Damit merkt sich ein BGP-Router, von welchem Nachbar-BGP-Router er welche BGP-Routing-Information erhalten hat. Allerdings werden in der Adj-RIB-IN nur die neuesten Routing-Informationen hinterlegt.

Diese Einträge bilden die Basis für die Bestimmung der besten Pfade von BGP. Aber bevor dies geschieht, werden auf alle neu eintreffenden Routing-Informationen die Policies angewendet. Das wird auch deutlich in Abbildung 2.3, wo die einzelnen Verarbeitungsschritte einer BGP-UPDATE-Nachricht in einem BGP-Router dargestellt sind.

Durch den Einsatz der Policies können unerwünschte Routing-Informationen gelöscht werden bzw. bei den in den BGP-UPDATE-Nachrichten enthaltenen Pfadvektoren können entsprechend der vordefinierten Regeln die Attribute verändert werden. Einen Überblick, welche Ziele man mit dem Konfigurieren der Attribute erreichen kann, wird in [CR05] gegeben.

Anhand dieser überarbeiteten BGP-Routing-Informationen werden die besten Pfade von BGP ausgewählt, wie es im Abschnitt 2.1.2.2 beschrieben wurde. Das Ergebnis dieser Pfadauswahl wird in der Loc-RIB abgelegt. Die Loc-RIB existiert nur einmal pro BGP-Router und ist der zentrale Speicher für die jeweils besten Pfade von BGP.

Aus der Loc-RIB heraus werden die Daten für die FIB generiert, anhand derer die Pakete weitergeleitet werden. Auf jeder Einsteckkarte eines Routers, die die Schnittstellen für die physikalischen Verbindungen bereitstellt, existiert eine FIB.

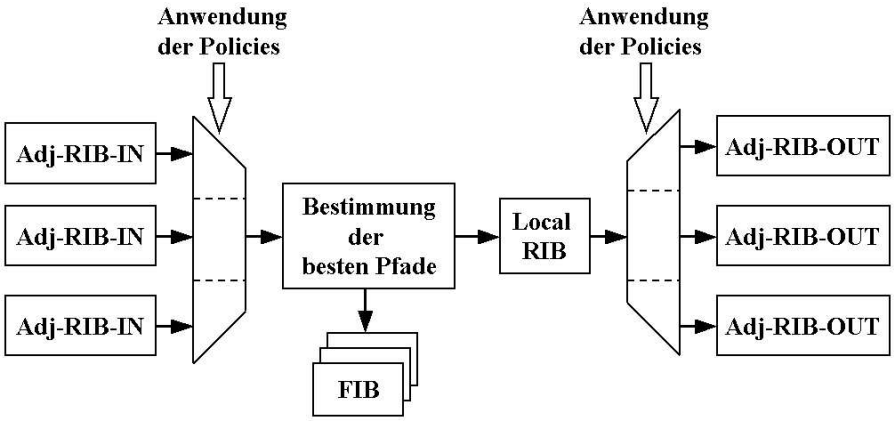


Abbildung 2.3: Bestimmung der besten Pfade in einem BGP-Router nach [Hal97]

Wurde in der Loc-RIB ein neuer bester Pfad abgelegt, wird dieser an die entsprechenden Nachbar-BGP-Router weitergeleitet. Bevor diese mittels einer BGP-UPDATE-Nachricht versendet werden, kommen die Policies nochmals zum Einsatz. Entsprechend dieser Regeln wird entschieden, an welche Nachbar-AS die BGP-UPDATE-Nachricht verschickt wird und welche Attribute des Pfadvektors noch verändert werden. Die darin enthaltenen Routing-Informationen werden wieder getrennt nach der jeweiligen BGP-Verbindung in die entsprechende Adj-RIB-OUT zwischengespeichert. Somit existiert, wie bei der Adj-RIB-IN auch, für jede BGP-Verbindung eine eigene Adj-RIB-OUT.

Mittels der Beeinflussung der BGP-Routing-Informationen beim Empfang und Versand einer BGP-UPDATE-Nachricht besitzt der Betreiber eines autonomen Systems eine wirkungsvolle Möglichkeit, um das BGP-Routing an seine Präferenzen anzupassen. Teilweise ist es dadurch auch möglich, Einfluß auf die BGP-Routing-Entscheidungen der benachbarten autonomen Systeme zu nehmen.

2.1.2.4 Einteilung der autonomen Systeme und deren Konsequenzen

Trotz der Freiheit, die jeder Betreiber beim Definieren der Policies hat, wurden in [Gao01] allgemein gültige Filterregeln in Abhängigkeit der Geschäftsbeziehung zwischen zwei AS-Betreibern aufgestellt. Dabei werden die Verbindungen in die drei Kategorien *Kunde-Betreiber*, *Peer-to-Peer* und *Sibling* eingeteilt.

Besteht zwischen zwei autonomen Systemen eine Kunde-Betreiber-Geschäftsbeziehung, so ist ein autonomes System der Kunde des anderen autonomen Systems. Für den Netzbetreiber des Kunden-AS entstehen Kosten für die Weiterleitung seiner Daten durch das Betreiber-AS. Um möglichst viel Verkehr zu erhalten, teilt das Betreiber-AS seinen Kunden-AS alle über ihn erreichbaren Präfixe mit. Somit erhöht er die Attraktivität, über ihn Verkehr zu übertragen, was wiederum zu gesteigerten Einkünften führen kann. Das Kunden-AS übermittelt dem Betreiber-AS nur die eigenen Präfixe und die Präfixe der eigenen Kunden-AS. Damit erhält das Betreiber-AS nur die Möglichkeit, Verkehr an das Kunden-AS und dessen Kunden-AS zu schicken.

Bei der Peer-to-Peer-Geschäftsbeziehung wird der Verkehr zwischen zwei autonomen Systemen ohne gegenseitige Bezahlung ausgetauscht. Für die beiden beteiligten autonomen Systeme ist dies nur sinnvoll, wenn in beiden Richtungen ungefähr die gleiche Datenmenge übertragen wird. Um diese Verbindungen nicht als preiswerte Verbindung zum Durchleiten des gesamten Verkehrs zu mißbrauchen, tauschen die beiden autonomen Systeme nur die eigenen Präfixe und die Präfixe ihrer Kunden aus. Damit wird verhindert, daß die Betreiber-AS die Peer-to-Peer-Verbindungen und die darüber erreichbaren Präfixe ihrer Kunden-AS kennen. Dasselbe gilt auch für die über eine Peer-to-Peer-Verbindung erreichbaren AS, an die keine Informationen über die Betreiber-AS übermittelt werden.

Interdomain-Verbindungen, die nicht diesen beiden Geschäftsbeziehungen zugeordnet werden können, werden unter Sibling-Geschäftsbeziehung eingeordnet. Dabei gibt es keine Einschränkungen zur Weitergabe von Präfixen oder sonstige Besonderheiten. Ein Beispiel für eine Sibling-Geschäftsbeziehung ist eine Interdomain-Verbindung zwischen zwei autonomen Systemen, die von demselben Netzbetreiber unterhalten werden.

Wie die eben beschriebenen Geschäftsbeziehungen die Weitergabe von Präfixen im allgemeinen einschränken, ist aus Sicht eines autonomen Systems in Tabelle 2.1 zusammengefaßt.

	Kunden-AS	Betreiber-AS	Peer-to-Peer-AS	Sibling-AS
Kunden-AS	x	x	x	x
Betreiber-AS	x	-	-	x
Peer-to-Peer-AS	x	-	-	x
Sibling-AS	x	x	x	x

Tabelle 2.1: Routing-Einschränkungen in Abhängigkeit der Geschäftsbeziehungen

Dabei stellt eine Spalte die Quelle der Interdomain-Verbindung und die Zeile die Senke dar. Ein x symbolisiert, daß von einem durch die Spalte typisierten autonomen System Interdomain-Verkehr zu dem durch die Zeile beschriebenen autonomen System geschickt werden kann.

Neben der Einordnung der Interdomain-Verbindungen anhand der Geschäftsbeziehungen der Netzbetreiber kann man die autonomen Systeme auch anhand ihrer Größe klassifizieren. In dieser Arbeit werden die autonomen Systeme in die drei Klassen *Tier1*, *Tier2* sowie *Tier3* eingeteilt. Als Tier1-AS oder *Transit*-AS bezeichnet man ein sehr großes autonomes System, welches unter Umständen die ganze Welt umspannen kann. Ein Transit-AS bietet als Service nur das Durchleiten von Datenverkehr an, der auch als Transitverkehr bezeichnet wird, d.h. ein Transit-AS generiert bzw. terminiert keinen Verkehr. Die Attraktivität eines Transit-AS ist seine Größe und die damit einhergehende Erreichbarkeit sehr vieler Ziele im Internet.

Zwischen allen Transit-AS im heutigen Internet besteht eine direkte physikalische Verbindung und somit sind diese untereinander vollvermascht. Da zwischen diesen Transit-AS ungefähr dieselbe Datenmenge ausgetauscht wird, sind diese Verbindungen alle Peer-to-Peer-Verbindungen. Über die restlichen Interdomain-Verbindungen eines Transit-AS sind die Kunden-AS angebunden, womit das Transit-AS hierbei als Betreiber fungiert.

Unter einem Tier2-AS versteht man ein autonomes System mittlerer Größe, das sich z.B. über ein ganzes Land oder über mehrere Staaten erstreckt. Die Tier2-AS besitzen normalerweise alle Interdomain-Verbindungstypen, und ihr Interdomain-Verkehr setzt sich aus Transit-Verkehr und aus in dem Tier2-AS generierten bzw. terminierten Verkehr zusammen.

Die kleinsten autonomen Systeme stellen die Tier3-AS oder *Stub*-AS dar. Mit Stub-AS ist gemeint, daß solch ein autonomes System nur Interdomain-Verkehr erzeugt oder terminiert und kein Verkehr durch dieses autonome System durchgeleitet wird. Somit tritt ein Stub-AS immer als Kunde eines größeren autonomen Systems auf.

2.1.2.5 Routen-Reflektoren

Die I-BGP-Kommunikation innerhalb eines autonomen Systems weist noch eine Besonderheit auf. Da allen BGP-Routern eines autonomen Systems die gleichen Routing-Informationen zur Verfügung stehen müssen, um Inkonsistenzen beim BGP-Routing zu vermeiden, tauschen alle BGP-Router eines autonomen Systems die besten Pfade untereinander aus. Dafür benötigt jeder BGP-Router eine BGP-Verbindung zu allen anderen BGP-Routern, die alle manuell in jedem BGP-Router konfiguriert werden müssen. Aufgrund dessen, daß jeder BGP-Router eine BGP-Verbindung zu allen anderen BGP-Routern besitzt, entsteht eine Vollvermaschung zwischen den BGP-Routern.

Besitzt ein autonomes System nun eine Vielzahl an BGP-Routern, dann ist die Konfiguration dieser Router sehr aufwendig. Wird ein neuer BGP-Router in ein Netz integriert, muß in diesem jeweils eine BGP-Verbindung zu allen anderen BGP-Routern manuell konfiguriert werden, und in den schon existierenden BGP-Routern muß jeweils eine BGP-Verbindung zu dem neuen BGP-Router aufgesetzt werden. Um diesen Aufwand zu verringern, wurden spezielle BGP-Router, die *Routen-Reflektoren* [BCC06], entwickelt. Dabei muß nur noch eine BGP-Verbindung von jedem BGP-Router zu einem Routen-Reflektor (RR) eingerichtet werden. Um weiterhin alle BGP-Router eines autonomen Systems mit denselben Routing-Informationen zu versorgen, sendet ein BGP-Router seine Routing-Information an den Routen-Reflektor und dieser verteilt diese an alle anderen BGP-Router.

Aus Performanz- und Verfügbarkeitsgründen können in einem autonomen System auch mehrere Routen-Reflektoren eingerichtet werden, wobei eine hierarchische Anordnung von Routen-Reflektoren ebenfalls vorgesehen ist. Der Vorteil des Einsatzes von Routen-Reflektoren soll im folgenden anhand von drei verschiedenen Topologien gezeigt werden.

Im ersten Beispiel wird ein kleines Netz mit nur drei BGP-Routern untersucht, das in Abbildung 2.4 (a) dargestellt ist.

Wird kein Routen-Reflektor benutzt, müssen die drei BGP-Router mit drei BGP-Verbindungen vollvermascht werden. Kommt ein Routen-Reflektor zum Einsatz, müssen ebenfalls drei BGP-Verbindungen konfiguriert werden, jeweils eine BGP-Verbindung von einem BGP-Router zu dem Routen-Reflektor (Knoten R_1 in Abbildung 2.4 (b)). Aufgrund der kleinen Router-Anzahl

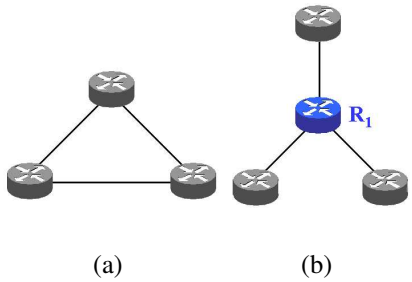


Abbildung 2.4: Netz mit drei BGP-Routern ohne (a) bzw. mit Routen-Reflektor (b)

ergibt sich in diesem Fall kein Vorteil. Erst wenn diese Topologie um weitere BGP-Router ergänzt wird, zahlt sich der Einsatz eines Routen-Reflektors durch einen geringeren Konfigurationsaufwand aus.

In dem zweiten Beispielnetz, das in Abbildung 2.5 zu sehen ist, bringt die Verwendung von Routen-Reflektoren deutliche Vorteile.

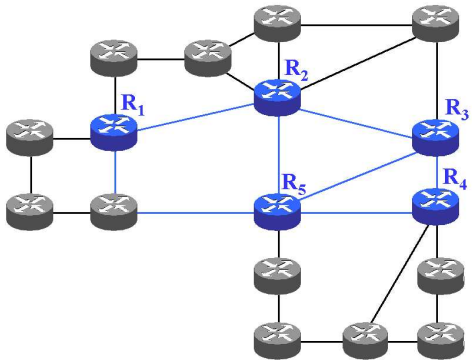


Abbildung 2.5: Netz mit siebzehn BGP-Routern

Dieses Netz wird in [Zus06] als NOBEL-Germany-Network beschrieben und besteht aus 17 Knoten und 26 bidirektionalen Verbindungen. Setzt man in diesem Netz keinen Routen-Reflektor ein, dann müssen $\frac{n*(n-1)}{2} = 136$ BGP-Verbindungen aufgesetzt werden. Wählt man die fünf BGP-Router mit dem größten Vermaschungsgrad als Routen-Reflektoren (R_1, R_2, R_3, R_4 und R_5 in Abbildung 2.5) aus, benötigt man für die Vollvermaschung zwischen den fünf Routen-Reflektoren zehn BGP-Verbindungen. Dazu müssen noch zwölf BGP-Verbindungen zwischen den restlichen BGP-Routern zu den jeweiligen Routen-Reflektoren addiert werden. Damit läßt sich die Anzahl der BGP-Verbindungen um 114 bzw. um 84% im Vergleich zur Konfiguration ohne Routen-Reflektoren reduzieren.

Mit 48 Knoten und 78 bidirektionalen Verbindungen ist das Beispielnetz aus Abbildung 2.6 das größte der drei hier betrachteten Topologien.

Setzt man keine Routen-Reflektoren in diesem Netz ein, benötigt man 1128 BGP-Verbindungen

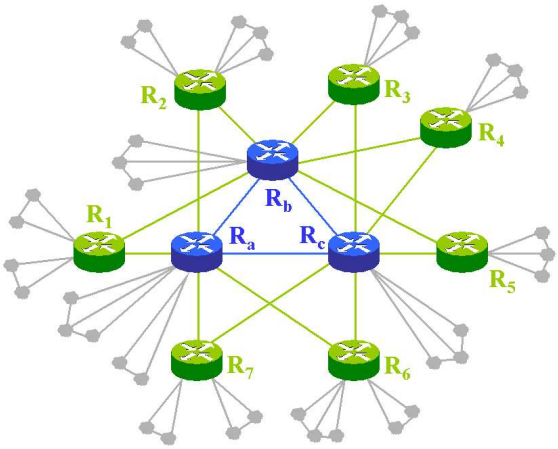


Abbildung 2.6: Netz mit 48 BGP-Routern

für die Vollvermaschung aller 48 BGP-Router. Wegen der Größe und der Struktur dieses Netzes werden im folgenden zwei unterschiedliche Konfigurationen mit Routen-Reflektoren untersucht. Beim ersten Aufbau bilden alle in Abbildung 2.6 farbig markierten BGP-Router (R_1 bis R_7, R_a, R_b, R_c) eine Ebene aus zehn Routen-Reflektoren. Für deren Vollvermaschung werden 45 BGP-Verbindungen benötigt. Die restlichen BGP-Router werden mit 38 BGP-Verbindungen mit den jeweiligen Routen-Reflektoren verbunden. Somit verringert sich die Anzahl der BGP-Verbindungen um 1045, was einer Einsparung von 93% entspricht.

Teilt man die zehn Routen-Reflektoren in zwei Ebenen auf, wird nur bei der oberen Hierarchiestufe eine Vollvermaschung notwendig. Als obere Ebene wurden die drei in Abbildung 2.6 blau markierten BGP-Router (R_a, R_b und R_c) ausgewählt, für deren Vollvermaschung sind drei BGP-Verbindungen nötig. Für die Anbindung der restlichen Routen-Reflektoren (R_1 bis R_7) an die oberste Ebene werden sieben weitere BGP-Verbindungen benötigt. Die Anzahl der BGP-Verbindungen der restlichen BGP-Router mit ihren jeweiligen Routen-Reflektoren ändert sich nicht. Daraus ergibt sich durch die Einführung einer zweiten Ebene eine Einsparung von 35 weiteren BGP-Verbindungen. Somit lassen sich im Vergleich zur Konfiguration ohne Routen-Reflektor 1080 BGP-Verbindungen einsparen, und die Reduktion vergrößert sich auf 96%.

Anhand dieser drei Beispielkonfigurationen wird deutlich, daß der Einsatz von Routen-Reflektoren eine deutliche Reduzierung des Konfigurationsaufwands bewirkt. Welche Einflüsse die Routen-Reflektoren auf die Fehlerreaktion von BGP haben, wird in Kapitel 3 ausführlich untersucht.

2.2 Garantie der Dienstqualität in IP-Netzen

Unter der Dienstqualität versteht man nach [IT93] eine Eigenschaft der Güte eines Dienstes, die durch die Zufriedenheit eines Nutzers beschrieben wird. Die Aufgabe eines Netzbetreibers besteht nun darin, die Nutzdaten der Dienste so zu übertragen, daß die von den Nutzern erwartete

Dienstqualität ohne wahrnehmbare Einbußen gewährleistet wird. Aber um dies sicherstellen zu können, werden Parameter benötigt, die die Nutzerzufriedenheit und damit die erwartete Dienstqualität beschreiben und die technisch überprüfbar sind.

2.2.1 Parameter zur Beschreibung der qualitativen Eigenschaften eines Dienstes

Für diesen Zweck wurden bei der IETF einige Parameter definiert, die diesen Anforderungen genügen. Ein sehr häufig genutzter Parameter, der die Güte der Datenübertragung eines Dienstes beschreibt, ist die verfügbare Kapazität einer Verbindung. Deren Maßeinheit ist in [PAM+98] mit Bit pro Sekunde definiert und beschreibt die Datenmenge, die eine Verbindung pro Zeiteinheit übertragen kann.

Desweiteren stellt die Verzögerung einen weiteren Parameter dar, der für die Beurteilung der Dienstqualität verwendet werden kann. Aber in diesem Zusammenhang steht die Verzögerung nicht für die reine Laufzeit einer Verbindung, sondern, wie in [AKZ99a] beschrieben, für die Zeit, die ein Bit benötigt, um von der Quelle zur Senke der Datenkommunikation zu gelangen.

Vor allem bei der Übertragung von Videodaten beeinträchtigen Paketverluste die Nutzerzufriedenheit. Deswegen wird dieser Parameter ebenfalls für die Beschreibung der Dienstqualität verwendet. In [AKZ99b] wird als Paketverlust das Ereignis, daß ein Paket abgeschickt wird, aber nicht bei der Senke ankommt, definiert. Für die Beurteilung der Güte der Datenübertragung ist nicht das Ereignis, sondern deren Häufigkeit das entscheidende Kriterium. Deswegen wird der Paketverlust auch als die Wahrscheinlichkeit, daß ein Paket verloren geht, beschrieben.

Ebenfalls für die Übertragung von Echtzeitdaten ist das Schwanken der Verzögerung von verschiedenen Paketen, die zu demselben Verkehrsfluß gehören, ein Effekt, der die Nutzerzufriedenheit beeinträchtigen kann. Dieser Effekt wird als *Jitter* bezeichnet und eignet sich ebenfalls als Meßgröße für die Überprüfung der Dienstqualität.

Ein Parameter, der häufig auch als eigenständige Eigenschaft der Datenübertragung betrachtet wird, ist die Verfügbarkeit. In dieser Arbeit wird die Verfügbarkeit ebenfalls als Parameter betrachtet, der die Güte einer Datenübertragung beschreibt. Kann die erwartete Dienstqualität nicht zu dem Zeitpunkt gewährleistet werden, an dem sie der Nutzer in Anspruch nehmen möchte, kann der Netzbetreiber die Erwartungshaltung des Nutzers nicht erfüllen.

Um die Verfügbarkeit überprüfen zu können, wird diese als Wahrscheinlichkeit beschrieben, mit der die zugesicherte Dienstqualität eingehalten werden kann. Dafür wird die Verfügbarkeit V aus dem Verhältnis der Zeit zwischen zwei Fehlern und der Summe aus dieser Zeit und der Reparaturzeit bestimmt:

$$\mathrm{V} = \frac{\mathrm{MTTF}}{\mathrm{MTTF} + \mathrm{MTTR}} \tag{2.1}$$

Die mittlere Zeit, in der ein fehlerfreier Betrieb möglich ist, wird als die Mean Time To Fail

(MTTF) bezeichnet, und die Mean Time To Repair (MTTR) beschreibt die mittlere Zeit, die für die Fehlerbehebung benötigt wird.

Da die Reparaturzeit ein entscheidender Faktor bei der Bestimmung der Verfügbarkeit ist, muß diese minimiert werden, um die Garantie der Dienstqualität aufrechterhalten zu können. Dafür ist eine schnelle Fehlerreaktion unabdingbar, weil diese der Reparaturzeit gleichgesetzt werden kann. Aus diesem Grund wird der Fehlerreaktion von BGP bei den Untersuchungen in Kapitel 3 und 4 besondere Beachtung geschenkt.

Zwischen der verfügbaren Kapazität und den Parametern Paketverlustwahrscheinlichkeit, Verzögerung sowie Jitter besteht eine Abhängigkeit. Reicht die Kapazität einer Verbindung nicht aus, dann werden die Pakete, die über diese Verbindung übertragen werden, in den Warteschlangen der betroffenen Router zwischengespeichert. Die daraus resultierende Wartezeit erhöht die Verzögerung der Pakete.

Sind die Warteschlangen aus Kostengründen und wegen der Begrenzung der maximalen Wartezeit relativ klein ausgelegt, erhöht sich die Paketverlustwahrscheinlichkeit, wenn die verfügbare Kapazität einer Verbindung nicht mehr ausreicht und deswegen nicht mehr alle Pakete zwischengespeichert werden können. Desweiteren führen Schwankungen bei der Menge der zwischengespeicherten Pakete zu einem erhöhten Jitter.

Dieselben Konsequenzen treten auf, wenn aufgrund eines Fehlers der betroffene Verkehr über andere Verbindungen übertragen wird und dort nicht ausreichend Kapazität zur Verfügung steht.

Die Konsequenz aus diesem Zusammenhang ist eine deutlich größere Bedeutung der verfügbaren Kapazität einer Verbindung für die Sicherstellung der Dienstqualität, da diese die anderen Parameter beeinflußt. Dies führt zu dem Bestreben der Netzbetreiber, immer ausreichend Kapazität in den Netzen vorzuhalten, um eine Beeinträchtigung der Dienstqualität durch unzureichende Kapazität der Verbindungen zu verhindern. Das wird noch verstärkt durch die relativ geringen Kosten für die Bereitstellung von ausreichend großen Verbindungskapazitäten und der sehr kleinen Paketverlustwahrscheinlichkeit in heutigen Kernnetzen.

Aus diesen Überlegungen ergibt sich als einfachste Lösung die Überdimensionierung, um die geforderte Dienstqualität zu gewährleisten. Allerdings ist dieser Weg aufgrund der sehr hohen Kapazitätsanforderungen sehr ineffizient, vor allem wenn man nicht nur den fehlerfreien Fall betrachtet, sondern auch nach einem Ausfall die zugesicherte Dienstqualität sicherstellen will.

2.2.2 Architekturen für die Garantie der Intradomain Dienstqualität

Für die Gewährleistung der Dienstqualität innerhalb eines IP-Netzes sowie für die Verwaltung und Reservierung der Ressourcen sind neben der Überdimensionierung viele Mechanismen entwickelt worden. Diese lassen sich in zwei grundsätzliche QoS-Architekturen einteilen. Beide wurden bei der IETF standardisiert und werden im folgenden kurz vorgestellt.

2.2.2.1 Integrated-Services-Architektur

Die *Integrated-Services* (IntServ)-Architektur [BCS94] wurde als erstes entwickelt und garantiert für jeden Verkehrsfluß die angeforderten Parameter, was auch als "harte" Garantie bezeichnet wird. Um dies zu ermöglichen, werden für jeden Verkehrsfluß entlang seines Pfads durch das Netz die benötigten Ressourcen reserviert. Dazu wird das *Resource Reservation Protocol* (RSVP) [BZB+97] verwendet. Um die Ressourcenreservierungen vorzunehmen, sendet RSVP eine spezielle Nachricht von der Quelle zur Senke, die abfragt, ob die benötigten Ressourcen auf dem Pfad verfügbar sind. Bei Erfolg werden die entsprechenden Ressourcen in jedem Router von einer RSVP-Nachricht auf dem Rückweg reserviert. Daraus resultiert, daß jeder Router in einem IntServ-Netz den jeweiligen Zustand jeder erfolgreichen Reservierung verwalten muß.

Neben RSVP sind noch weitere Komponenten für die Gewährleistung der Dienstqualität notwendig, die ebenfalls in [BCS94] definiert sind. Dazu gehören ein Netzzugangskontrollverfahren, der *Scheduler* für die Verarbeitung der Pakete in einem Router und ein Mechanismus für die Markierung der Pakete. Das Netzzugangskontrollverfahren stellt sicher, daß nur die Menge an QoS-Verkehr in das Netz gelangt, für die auch Ressourcen reserviert wurden. Für die Entscheidung am Rand des IntServ-Netzes, ob ein QoS-Verkehrsfluß in das Netz weitergeleitet wird oder nicht, ist eine Schnittstelle zu RSVP notwendig, um eine Ressourcenreservierung initiieren zu können.

Für die Erkennung, ob es sich um QoS-Verkehr handelt, muß jedes Paket eine Markierung tragen. Anhand dieser Markierung erkennt ein Router auch, wie sein Scheduler das entsprechende Paket behandeln muß, um die Dienstqualität sicherzustellen. Da nicht für jeden Verkehrsfluß eigene Regeln für die Verarbeitung der Pakete aufgestellt werden können, werden in [Wro97, SPG97] zwei verschiedene Klassen für den QoS-Verkehr inklusive der Verarbeitungsroutinen für den Scheduler definiert. Aufgrund der unterschiedlichen Markierung der einzelnen Pakete kann der Scheduler eines Routers entscheiden, welcher Klasse diese Pakete zuzuordnen sind, und kann dementsprechend einen der beiden Verarbeitungsprozesse auswählen.

Mittels dieser vier Komponenten kann ein Netzbetreiber mit der IntServ-Architektur für jeden Verkehrsfluß die Dienstqualität innerhalb eines Netzes garantieren. Da aber jeder Router den Zustand der entsprechenden Verkehrsflüsse verwalten muß, skaliert die IntServ-Architektur nicht besonders gut. Trotz der Versuche, dieses Skalierbarkeitsproblem durch Verbesserungen bei RSVP und Reduzierung der Leistungsanforderungen an die Router [BGS+01, BILF+01] zu beheben, eignet sich die IntServ-Architektur nicht für große Netze mit sehr vielen Verkehrsflüssen. Damit ist sie auch nicht geeignet, um als Basis für die Gewährleistung der Interdomain-Dienstqualität verwendet zu werden.

2.2.2.2 Differentiated-Services-Architektur

Um das Skalierungsproblem der IntServ-Architektur zu vermeiden, wurde die *Differentiated-Services* (DiffServ)-Architektur [BBC+98] entwickelt. Die Grundidee besteht darin, den Verkehr in verschiedene Klassen einzuteilen und alle Verwaltungsaufgaben für den Verkehr nur am Netzrand durchzuführen.

Am Eingangs-Router eines DiffServ-Netzes werden die eingehenden Pakete anhand ihrer Markierung einer Klasse von DiffServ – der sogenannten QoS-Klasse – zugeordnet. Als Markierung für die QoS-Klasse wird ein Byte im Header jedes IP-Paketes benutzt. Dieser Teil wird als *DiffServ Code Point* (DSCP) [NBB+98] bezeichnet.

Der DSCP dient innerhalb des Netzes dazu, die Pakete entsprechend den Anforderungen der QoS-Klasse weiterzuleiten. Wie dies in einem Router realisiert wird, ist im *Per Hop Behavior* (PHB) [NC01] beschrieben. Dabei wird bei DiffServ zwischen drei verschiedenen PHB unterschieden. Beim PHB *Expedited Forwarding* (EF) [DCB+02] wird der QoS-Verkehr ohne Rücksicht auf den restlichen Verkehr mit der höchsten Priorität verarbeitet. Eine Stufe darunter rangiert das PHB *Assured Forwarding* (AF) [HBW+99]. Der QoS-Verkehr, der diesem PHB zugeordnet wird, muß dem QoS-Verkehr mit der Markierung für das PHB EF Vorrang gewähren. Der heutige IP-Verkehr wird mit dem PHB *Best Effort* (BE) verarbeitet und wird ohne Zusage einer Dienstqualität sowie mit der niedrigsten Priorität weitergeleitet. Welche QoS-Klassen zu welchem PHB zugeordnet werden, kann vom Netzbetreiber konfiguriert werden und beeinflußt wesentlich die Netzplanung des entsprechenden Netzes.

Mit dem DiffServ-Ansatz wurde eine Architektur definiert, die sehr gut skaliert, da die wesentlichen Verarbeitungsschritte, wie das Einordnen des Verkehrs in die jeweiligen QoS-Klassen und die Entscheidung, ob die geforderte Dienstqualität zugesichert werden kann, ausschließlich am Netzrand erfolgt. Allerdings kann die Dienstqualität nur im Rahmen der Parameter der QoS-Klasse und dem damit verknüpften PHB gewährleistet werden. Deswegen spricht man bei DiffServ auch von einer "weichen" Garantie der Dienstqualität.

Der Ansatz, den QoS-Verkehr in QoS-Klassen einzuteilen, bietet eine sehr gute Skalierungsfähigkeit und wird deswegen auch für die Garantie der Interdomain-Dienstqualität weiterverwendet.

2.2.3 Stand der Technik und Ausblick

Die zwei grundsätzlichen Ansätze der IntServ- und DiffServ-Architektur werden in vielen Lösungen verwendet und sind dabei auch weiterentwickelt worden. In [BFY+00] wird sogar ein Ansatz vorgeschlagen, der die beiden Architekturen miteinander verknüpft und somit die Vorteile beider Ansätze nutzen kann. Desweiteren existieren viele Lösungen, die auf der DiffServ-Architektur aufsetzen, wie z.B. die Architekturen in [EGH+03, Myk03, Gla03, Tro05]. Neben der Entwicklung dieser Architekturen wurde viel Wert darauf gelegt, die Gewährleistung der Dienstqualität in die Netzplanung mit einfließen zu lassen. Zum Teil sind bei diesen Arbeiten auch neue Verfahren für die Netzplanung von QoS-Verkehr entstanden, wie z.B. [Sha05b, Men04].

Zusammenfassend kann man feststellen, daß für die Garantie der Dienstqualität von QoS-Verkehr innerhalb eines Netzes viele ausgereifte Lösungen existieren und diese teilweise auch schon als Produkte angeboten werden. Allerdings ist dieses Angebot nur auf die Gewährleistung der Intradomain-Dienstqualität beschränkt. Für die Garantie der Interdomain-Dienstqualität existieren erste Ansätze, die im Abschnitt 4.2 näher untersucht werden. Da diese den Anforderungen

nur teilweise gerecht werden, wird in Kapitel 4 ein neues Konzept für das Sicherstellen der Interdomain-Dienstqualität entwickelt.

2.3 Netzplanung

Wie im letzten Abschnitt erläutert wurde, ist es für die Garantie der Dienstqualität erforderlich, ausreichend Ressourcen in einem Netz bereitzustellen. Der dahinterstehende Prozeß ist die Netzplanung, die jeder Netzbetreiber in verschiedenen Situationen ausführt. Neben der Auslegung des Netzes im Vorfeld des Netzaufbaus ist die Durchführung der Netzplanung auch beim etappenweisen Ausbau des Netzes oder bei der Anpassung an neue Gegebenheiten notwendig.

Der Netzplanungsprozeß wird in Kapitel 5 anhand eines Beispiels durchlaufen, um die Einsparungen des Kapazitätsbedarfs des neuen Dimensionierungsverfahrens, das die Anforderungen des Interdomain-QoS-Verkehrs erfüllt, zu belegen. Deswegen wird im folgenden kurz auf den grundlegenden Ablauf der Netzplanung eingegangen.

Die Auslegung eines einzelnes Netz basierend auf den Anforderungen an dieses Netz läßt sich in mehrere Einzelschritte unterteilen, deren zeitlicher Ablauf in Abbildung 2.7 dargestellt ist.

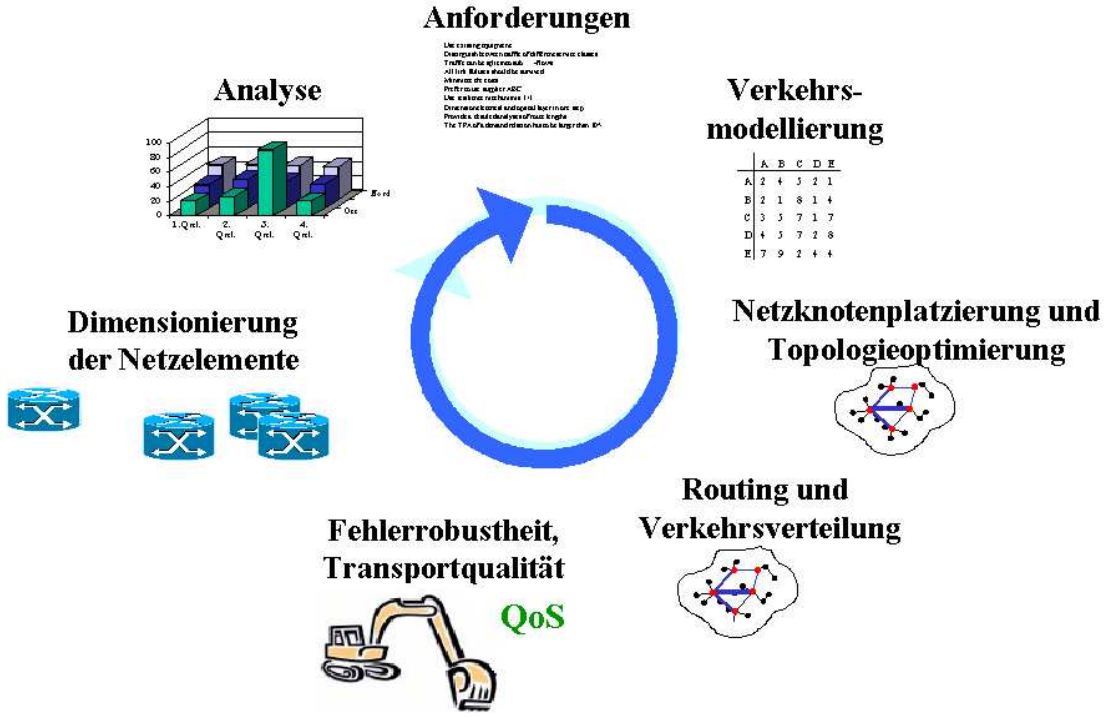


Abbildung 2.7: Netzplanungszyklus

Generell werden alle Schritte des gesamten Kreislaufs immer ausgeführt. Allerdings kann es durchaus vorkommen, daß die vorgegebenen Randbedingungen dazu führen, daß einzelne Schritte übersprungen werden. Soll z.B. ein bestehendes Netz an geänderte Verkehrsanforderungen

angepaßt werden, erübrigt sich die Netzknotenplazierung und Topologieoptimierung, wenn die vorhandene Netztopologie beibehalten werden soll.

Das Ziel der Netzplanung ist es, ein Netz so zu dimensionieren, daß der zu übertragende Verkehr möglichst kostengünstig im Netz transportiert werden kann. Dabei müssen die zu betrachtenden Kosten nicht ausschließlich die monetären Kosten des Netzes sein.

Der Ausgangspunkt der Netzplanung sind die Anforderungen an das Netz. Diese leiten sich aus den Kundenanforderungen und den Randbedingungen des Betreibers ab. Die Leistungen von bzw. für benachbarte Netze werden in Verträgen, den so genannten *Service Level Agreement* (SLA), festgelegt.

Ein SLA besteht jeweils zwischen einem Kunden und dem Netzbetreiber, dessen Netz geplant werden soll. Dabei geht es darum festzuschreiben, welche Dienste mit welcher Güte und zu welchem Preis der Kunde nutzen kann. Auf der anderen Seite legt der Netzbetreiber mit einem SLA mit einem anderen Netzbetreibern fest, welche Menge an Verkehr zu welchen Bedingungen zwischen diesen beiden Netzbetreibern ausgetauscht werden kann. Eine Auswahl an möglichen Parametern, die in einem SLA festgelegt werden können, beschreibt die folgende Aufzählung, wobei diese Liste beliebig fortgesetzt werden kann.

- Dienstqualität des Verkehrs (QoS-Klasse, benötigte Kapazität, Paketverlustwahrscheinlichkeit, Verzögerung etc.)
- Verfügbarkeit, Fehlerhäufigkeit
- Sicherheitsvorkehrungen gegen den Mißbrauch und Angriffe Dritter
- Meßmethoden zur Überprüfung der zugesicherten Leistung
- Kosten der angebotenen Leistung
- Vertragsstrafen bei Nichteinhaltung der zugesagten Leistung

Neben diesen vertraglichen Anforderungen gibt auch der Betreiber weitere zu beachtende Randbedingungen vor. Neben der wichtigsten Bedingung, der Minimierung der Investitions- und Betriebskosten (CAPEX und OPEX), können auch noch weitere Punkte von Bedeutung sein. Mögliche Beispiele sind die Wiederverwendung von schon vorhandenen Standorten, Einschränkung der zu verwendenden Technologien, Designvorgaben für die Topologie sowie die Forderung einer bestimmten Ausfallsicherheit.

Nach der Festlegung der zu erfüllenden Anforderungen ist der nächste Schritt die Modellierung des Verkehrs. Dabei kommt es darauf an, den zukünftigen Verkehr, für den das Netz geplant werden soll, festzulegen. Dies kann auf Messungen im aktuellen Netz und auf Erfahrungswerten über das Wachstum des Verkehrs beruhen. Wenn diese Daten nicht vorliegen, können u.a. die Bevölkerungsstrukturen als Basis für die Vorhersage berücksichtigt werden. Neben der Bestimmung der Menge des Verkehrs müssen auch weitere Eigenschaften festgelegt werden, wie z.B.

die Einteilung in QoS-Klassen. Aus diesen Daten muß letztendlich eine Verkehrsmatrix gebildet werden. Die Elemente dieser Verkehrsmatrix stellen die Verkehrsmenge, die von den Quellen (Spalten der Matrix) zu den entsprechenden Senken (Zeilen der Matrix) übertragen werden soll, dar.

Nach der Erstellung der Verkehrsmatrix müssen im nächsten Schritt die Knoten plaziert werden. Dabei gilt es, eventuell vorhandene Standorte zu berücksichtigen und zu entscheiden, an welchen Standorten Verkehr mit anderen Netzen ausgetauscht, Zugangsnetze und besondere Kunden angebunden werden sollen. Wenn die Knoten plaziert wurden, müssen diese miteinander verbunden werden. Wichtig ist dabei, die Anforderungen der Fehlerrobustheit zu beachten, die z.B. redundante Verbindungen für das Ersatzschalten im Fehlerfall vorschreiben.

Basierend auf der Verkehrsmatrix und der Topologie kann das Routing des Verkehrs ermittelt werden. Welche Routing-Strategien angewendet werden, hängt stark von den Anforderungen des jeweiligen Netzbetreibers ab. Die bekanntesten IP-Routing-Protokolle wurden in Abschnitt 2.1.1 vorgestellt. Aus der Bestimmung des Routing resultiert die Verteilung der einzelnen Verkehrsflüsse auf die Verbindungen des Netzes und damit sind die Kapazitätsanforderungen der Verbindungen vorläufig festgelegt.

Im Anschluß daran werden die Mechanismen für die Fehlertoleranz und für die Zusicherung der Dienstqualität geplant. Bei der Fehlertoleranz gilt es abzusichern, daß auch im Falle der betrachteten Fehlerbilder der Verkehr ohne Einbußen weiter transportiert werden kann. Sofern die Dienstqualität nicht schon bei der Bestimmung der Verkehrsmatrix berücksichtigt wurde, müssen die dafür notwendigen Mechanismen nun mit eingeplant werden. Daraus und aus den Anforderungen der Fehlerrobustheit ergeben sich weitere Kapazitätsanforderungen, die zu den bisherigen hinzugezählt werden müssen.

Mit diesem Wissen können die Netzkomponenten, wie die physikalischen Verbindungen und die Netzknoten, dimensioniert werden. Dabei werden die Menge der benötigten Komponenten, deren Leistungsfähigkeit, usw. festgelegt.

Als letzter Schritt wird überprüft, ob das dimensionierte Netz den gestellten Anforderungen auch genügt. Ist dies der Fall, ist der Netzplanungszyklus abgeschlossen. Sind einige Anforderungen nicht erfüllt, müssen einzelne Schritte oder der komplette Prozeß wiederholt werden. Um zu vermeiden, daß erst am Zyklusende der Netzplanung die Nichteinhaltung der Anforderungen festgestellt wird, ist es sinnvoll, schon nach jedem einzelnen Schritt zu kontrollieren, ob alle Bedingungen erfüllt sind. Die so erweiterte Netzplanung ist in Abbildung 2.8 dargestellt.

Der Vorteil der Überprüfung der Bedingungen nach jedem Schritt ist, daß man sofort die entsprechenden Schritte wiederholen kann. Die roten Pfeile in Abbildung 2.8 symbolisieren die nach jeder Analyse möglichen und sinnvollen Wiederholungen.

Die Herausforderung bei der Netzplanung ist die gleichzeitige Betrachtung aller Einzelschritte, was zum bestmöglichen Ergebnis führen würde. Allerdings ist dies aus Komplexitätsgründen nicht möglich, wobei mehrere Planungsschritte gemeinsam zu optimieren, heute Stand der Technik ist. Je nach Planungsaufgabe, kann man verschiedene Schritte miteinander kombinieren. Ein Beispiel, das zu deutlichen Verbesserungen führt, ist die gleichzeitige Optimierung des Routing

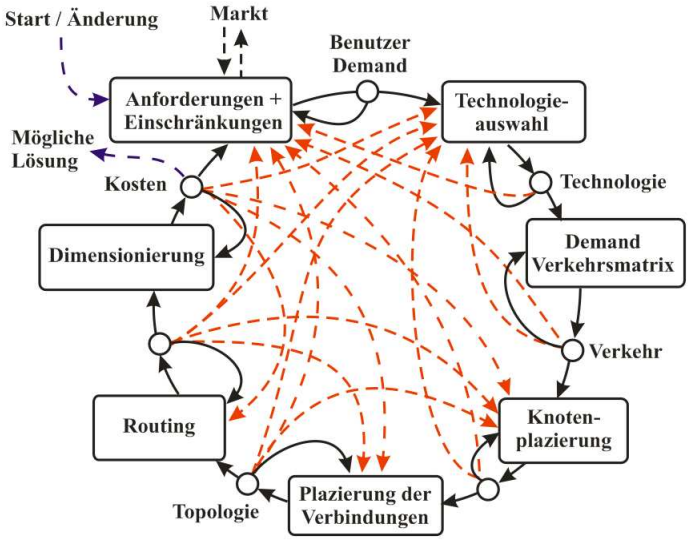


Abbildung 2.8: Detaillierter Netzplanungszyklus

in Kombination mit verschiedenen Fehlertoleranzmechanismen und der Dimensionierung der Netzknoten und Verbindungen.

2.4 Internet-Topologien

Die Größe des Internets mit mehr als 20.000 autonomen Systemen verhindert Untersuchungen, die auf die Protokollabläufe auf Paket-Ebene im gesamten Internet angewiesen sind. In [Amm05] wird anschaulich aufgezeigt, daß davon auszugehen ist, daß dies auch in Zukunft nicht möglich sein wird.

Diese Einschränkungen müssen auch bei den Simulationen der Fehlerreaktion von BGP in Kapitel 3 berücksichtigt werden. Für diese Untersuchungen können Internet-Topologien mit maximal 400 autonomen Systemen verwendet werden, da bei einer größeren Anzahl an autonomen Systemen die Grenze von 2 GB Arbeitsspeicher pro Prozeß überschritten wird. Diese maximal mögliche Anzahl an autonomen Systemen läßt sich durch den Einsatz von anderer Hardware bis zur Grenze des Simulators bei weniger als 800 autonomen Systemen noch erhöhen, aber die untersuchbaren Topologien bleiben wesentlich kleiner als die reale Größe des Internets.

Ein zweiter wichtiger Grund, der die Verwendung der realen Internet-Topologie verhindert, ist die Freiheit der Netzbetreiber der autonomen Systeme bei der Konfiguration des BGP-Routing. Die durch die Policies beschriebene Beeinflussung der Suche der besten Pfade von BGP führt dazu, daß auf IP-Ebene keine globalen Informationen über die reale Internet-Topologie existieren. Da wegen der Policies jedes autonome System unterschiedliche Routing-Informationen besitzt, die nicht alle möglichen Pfade enthalten, entwickelt jedes autonome System seine eigene unvollständige Sichtweise auf die reale Internet-Topologie.

Diese beiden Tatsachen führen dazu, daß man künstliche Internet-Topologien, die die Untersuchung des Zeitverhaltens von BGP ermöglichen, benötigt. Die Generierung von Topologien bietet den Vorteil, daß man Topologien mit bestimmten Eigenschaften, wie z.B. mit vorher festgelegter Anzahl von Knoten, erstellen kann. Welche Algorithmen für das Erzeugen von Internet-Topologien am besten geeignet sind und wie die in dieser Arbeit verwendeten Internet-Topologien erzeugt wurden, wird im folgenden beschrieben.

2.4.1 Eigenschaften der Internet-Topologie

Bevor darauf eingegangen wird, wie man eine Internet-Topologie erzeugen kann, muß bekannt sein, welche Eigenschaften eine Internet-Topologie besitzt und wodurch sie sich von anderen Topologien unterscheidet. Unabhängig vom steten Wachstum der realen Internet-Topologie weisen alle Internet-Topologien aus verschiedenen Wachstumsstadien einige wesentliche Gemeinsamkeiten auf.

In [FFF99] sind anhand von Untersuchungen und Messungen drei Zusammenhänge, die alle der Potenzgesetzverteilung unterliegen, ermittelt worden. Der Ausgangspunkt des ersten Zusammenhangs ist die Reihenfolge der sortierten autonomen Systeme anhand ihres Vermaschungsgrads. Die Plazierung $\mathrm{r_v}$ eines autonomen Systems in dieser Reihenfolge potenziert mit der Konstante R ist proportional zu dem Vermaschungsgrad $\mathrm{d_v}$ des autonomen Systems v. Diese Proportionalität beschreibt auch die folgende Gleichung:

$$\mathrm{d_v} \propto \mathrm{r_v^R} \tag{2.2}$$

Als Proportionalitätsfaktor wurde in [FFF99] der Kehrwert der Anzahl der autonomen Systeme potenziert mit der Konstanten $\mathrm{R} \approx -0,8$ ermittelt.

Ein weiterer Zusammenhang besteht zwischen der Häufigkeit $\mathrm{f_d}$ eines Vermaschungsgrads und dem Vermaschungsgrad d potenziert mit der Konstante $\mathrm{Q} \approx -2,2$, wie es in Gleichung 2.3 beschrieben ist.

$$\mathrm{f_d} \propto \mathrm{d^Q} \tag{2.3}$$

Das dritte Potenzgesetz beschreibt die Charakteristika der Eigenwerten der Adjazenz-Matrix. Als Adjazenz-Matrix bezeichnet man die Darstellung, welche Knoten miteinander verbunden sind, in Form einer Matrix. Die Eigenwerte dieser Matrix werden häufig für die Charakterisierung von Graphen eingesetzt [MP02]. Faloutsos e.t. [FFF99] haben herausgefunden, daß bei der Internet-Topologie die ersten zwanzig Eigenwerte λ_i der Adjazenz-Matrix proportional zur Potenz der Ordnung i mit der Konstante $\mathrm{E} \approx -0,5$ sind.

$$\lambda_\mathrm{i} \propto \mathrm{i^E} \tag{2.4}$$

Neben diesen drei Eigenschaften der realen Internet-Topologie wird in [BT02] gezeigt, daß die Internet-Topologie auch den Gesetzmäßigkeiten eines *Small World Graph* [WS98] folgt. Ein Small World Graph ist dadurch gekennzeichnet, daß der *clustering*-Koeffizient wesentlich größer ist als der eines Zufallsgraphen, der die selbe charakteristische Pfadlänge aufweist. Als charakteristische Pfadlänge bezeichnet man die mittlere Anzahl an Knoten eines Netzes, die ein Pfad durchläuft, um sein Ziel zu erreichen. Der Clustering-Koeffizient C eines Netzes berechnet sich nach [WS98] als Mittelwert über alle Knoten des Netzes des Verhältnisses $\mathrm{C_v}$ der Anzahl der Verbindungen $\mathrm{k_v}$ eines Knoten v zu der Verbindungsanzahl eines vollvermaschten Netzes. Diesen Zusammenhang beschreibt Gleichung 2.5, wobei n die Anzahl aller Knoten des Netzes darstellt.

$$\mathrm{C} = \frac{1}{\mathrm{n}} * \sum_{\mathrm{v}=1}^{\mathrm{n}} \mathrm{C_v} = \frac{1}{\mathrm{n}} * \sum_{\mathrm{v}=1}^{\mathrm{n}} \frac{\mathrm{k_v}}{\mathrm{k_v} * (\mathrm{k_v} - 1)/2} \tag{2.5}$$

Für den Nachweis, daß die Internet-Topologie einem Small World Graph entspricht, wurde in [BT02] die reale Internet-Topologie in einem Zeitraum von September 1999 bis Januar 2001 basierend auf den öffentlich verfügbaren BGP-Routing-Tabellen des *Oregon Route Views*-Projekts [rou05] und des *Routing Information Service* von Réseaux IP Européens (RIPE) [rib06] untersucht. In diesem Zeitraum ist das Internet um mehr als 100% gewachsen, und die charakteristische Pfadlänge sowie der Clustering-Koeffizient dieser Internet-Topologien haben sich nur um 2% bzw. 18% geändert. Mit einem ermittelten Clustering-Koeffizient von ungefähr 0,46 bei einer charakteristischen Pfadlänge von rund 3,6 wurde die Vermutung, daß die Internet-Topologie ein Small World Graph belegt, da der Clustering-Koeffizient eines Zufallsgraphen mit durchschnittlich 0,002 wesentlich kleiner ist.

Basierend auf diesen graphentheoretischen Eigenschaften kann bewertet werden, welche Algorithmen für die Generierung von Internet-Topologien geeignet sind.

2.4.2 Generierung von Internet-Topologien

Aus den verfügbaren Topologiegeneratoren gilt es denjenigen auszuwählen, der Topologien erzeugt, die den im letzten Abschnitt beschriebenen Eigenschaften der realen Internet-Topologie am nächsten kommen. In [BT02] werden die bekanntesten Topologiegeneratoren für Internet-Topologien miteinander verglichen. Als Metrik wurden dafür nur der Zusammenhang aus der Gleichung 2.2 und die beiden Eigenschaften des Small World Graph benutzt. In [MMB00] wurde gezeigt, daß die ersten beiden Zusammenhänge aus [FFF99] (beschrieben in den Gleichungen 2.2 und 2.3) hinreichend sind und der dritte Zusammenhang aus [FFF99] auch für andere Graphen gilt. Da die Gleichungen 2.2 und 2.3 jeweils verschiedene Zusammenhänge in Abhängigkeit des Vermaschungsgrads beschreiben, wurde in [BT02] nur der zweite berücksichtigt. Allerdings wurde dieser nicht exakt übernommen, sondern die empirische umgekehrte Verteilungsdichtefunktion, wie sie Gleichung 2.6 beschreibt, ist als Metrik herangezogen worden.

$$\mathrm{F(d)} = \sum_{\mathrm{i=d}}^{\infty} \mathrm{f(d)} \tag{2.6}$$

$$\mathrm{F(d)} \propto \mathrm{d}^{\alpha} \tag{2.7}$$

Die Proportionalität bleibt durch diesen Schritt erhalten. Aber der Exponent ändert sich und beträgt $\alpha = -1,155$.

Die besten Ergebnisse des Vergleichs in [BT02] liefert der *Generalized Linear Preference* (GLP)-Algorithmus, womit er sich gegen den *Power Law Random Graph*-Algorithmus [ACL00], den Generator nach Barabasi (BA) [BA99], den Generator nach Albert und Barabasi (AB) [AB00] und den *Internet Topology Generator* [JCJ00] durchsetzen konnte. Bestätigt werden diese Ergebnisse in [JRT04], wobei hier eine neue Bewertungsmetrik für die Internet-Ähnlichkeit eingeführt wurde.

Der GLP-Algorithmus baut auf dem BA-Generator auf, der ausgehend von einem kleinen Kernnetz die gewünschte Topologie entwickelt. Dazu fügt er bei jedem Schritt entweder einen neuen Knoten mit einer bestimmten Anzahl an Verbindungen oder nur diese Verbindungen zu der bestehenden Topologie hinzu. In Abhängigkeit der Wahrscheinlichkeit, mit der einer dieser beiden Schritte ausgeführt wird und wie die Verbindungen gelegt werden, entsteht eine Internet-Topologie, die der realen Internet-Topologie relativ ähnlich ist.

Diesen prinzipiellen Ablauf behält der GLP-Algorithmus bei. Allerdings ändert er die Wahrscheinlichkeit, zwischen welchen Knoten neue Verbindungen aufgebaut werden. Dabei wird die Eigenschaft, wie es in [CCG+02] beschrieben ist, berücksichtigt, daß eine neue Verbindung mit einer größeren Wahrscheinlichkeit zu einem Knoten mit einem höheren Vermaschungsgrad aufgebaut wird als umgekehrt. Aufgrund dieser Änderung liegt der Exponent Q aus der Gleichung 2.3 wesentlich näher an dem Idealwert, wie ihn die realen Internet-Topologie aufweist, und somit konnte ein großer Nachteil des BA-Generators behoben werden. In [BT02] konnte außerdem gezeigt werden, daß die Erweiterungen des AB-Generators keine wesentliche Verbesserung gegenüber dem BA-Generator darstellen und deshalb im GLP-Algorithmus nicht berücksichtigt wurden.

Somit empfiehlt es sich, den GLP-Algorithmus zum Erzeugen von Internet-Topologien für die Untersuchungen in Kapitel 3 und 4 zu verwenden.

2.4.3 Verwendete Internet-Topologien

Um den GLP-Algorithmus für die Generierung der in dieser Arbeit verwendeten Internet-Topologien einzusetzen, wurde auf den Topologiegenerator *Boston University Representative Internet Topology Generator* (Brite) [MLM+01] zurückgegriffen. Dieser bietet neben dem GLP-Algorithmus auch noch andere Algorithmen und eine Vielzahl an Einstellungsmöglichkeiten an. Für diese Arbeit wurden mittels Brite Internet-Topologien mit 20, 50, 100, 200 und 500 Knoten

erzeugt, wobei die Knoten die autonomen Systeme repräsentieren, die zur Vereinfachung nur aus einem BGP-Router bestehen. Dabei können die Eigenschaften dieser Topologien noch durch drei Parameter des GLP-Algorithmus – $\mathrm{p}, \beta, \mathrm{m}$ – beeinflußt werden.

Zum einen beschreibt p die Wahrscheinlichkeit, mit der in jedem Schritt ein neuer Knoten zu der erzeugten Topologie hinzugefügt wird. Das Komplementärereignis ist die Hinzunahme einer weiteren Verbindung. Mittels des Parameters β läßt sich die Wichtung, mit der ein neuer Knoten mit einem größeren Knoten verbunden wird, verstärken bzw. abschwächen. Als dritten Parameter kann man die Mindestanzahl m an Verbindungen, die ein Knoten besitzen soll, einstellen. Mit diesen drei Parametern läßt sich der Parameter α berechnen, wie in Gleichung 2.8 beschrieben ist.

$$-\alpha = \frac{2 * \mathrm{m} - \beta * (1 - \mathrm{p})}{(1 + \mathrm{p}) * \mathrm{m}} \tag{2.8}$$

Dieser Parameter α entspricht dem Exponenten der Potenzgesetzverteilung, wie sie in [BT02] aufgestellt wurde und in Gleichung 2.3 zusammengefaßt ist.

Da die Topologien für die Untersuchungen der Fehlerreaktion von BGP benutzt werden und dabei der Zerfall der Topologie in zwei Teile nach einem Fehler vermieden werden muß, wurde die minimale Verbindungsanzahl eines Knoten $\mathrm{m} = 2$ gewählt. Um dem in [BT02] ermittelten Wert von $\alpha = -1,155$ so nahe wie möglich zu kommen, wurden die beiden anderen Parameter anhand der Gleichung 2.8 mit $\mathrm{p} = 0,45$ und $\beta = 0,9$ ermittelt. Somit ergibt sich ein Exponent von $\alpha = -1,208$. Verglichen mit den in [BT02] untersuchten Internet-Topologien ergeben sich somit Internet-Topologien, die dem Ideal der realen Internet-Topologie noch besser entsprechen.

Aufgrund der Eigenschaften der erzeugten Internet-Topologien können von den Untersuchungen Rückschlüsse auf das Verhalten im realen Internet gezogen werden. Bei der Beurteilung des Konvergenzverhaltens von BGP ist es besonders wichtig, daß die Internet-Topologie einen Small World Graph darstellt, und somit der Mittelwert der kürzesten Pfade gemessen in AS-Hops der gleiche ist wie in der Realität. Damit ist sichergestellt, daß die Ausbreitung der BGP-Nachrichten sich sehr ähnlich verhält.

Ein wesentlich größerer Unterschied zwischen der realen Internet-Topologie und den generierten Topologien besteht bei der Anzahl der zu verwaltenden Präfixe. Das spiegelt sich besonders beim Übertragen aller Informationen der BGP-Routing-Tabelle und somit auch bei der Anzahl der versendeten BGP-Nachrichten wider. Betrachtet man aber nur die Anzahl der verschickten BGP-Nachrichten pro Präfix, so ist aufgrund des Small World Graph eine sehr ähnliche Anzahl zu erwarten.

Damit bilden die Internet-Topologien eine wichtige Grundlage für die Untersuchungen der folgenden Kapitel, und es ist dabei sichergestellt, daß die Eigenheiten der realen Internet-Topologie berücksichtigt werden.

2.5 Zusammenfassung

In diesem Kapitel wurden die grundlegenden Protokolle und Verfahren, die für die folgenden Untersuchungen relevant sind, beschrieben. Der erste Teil enthält einen Überblick über die verschiedenen Intra- und Interdomain-Routing-Protokolle in IP-Netzen. Dabei wurde näher auf das Interdomain-Routing-Protokoll BGP eingegangen. Neben dem prinzipiellen Ablaufs dieses Protokolls wurde ausführlich die Bedeutung und die Einflußnahme der Policies sowie die Funktionsweise der Routen-Reflektoren erläutert. Diese Kenntnisse werden benötigt, um in Kapitel 3 das heutige Interdomain-Routing zu erweitern, so daß die Konvergenzzeit minimiert wird und trotzdem die Kompatibilität zu BGP gewährleistet bleibt. Desweiteren ist das Interdomain-Routing eine wichtige Einflußgröße für die Untersuchungen in den Kapiteln 4 und 5.

Um das Konzept für die Zusicherung der Interdomain-Dienstqualität entwickeln zu können, ist es notwendig, die Mechanismen für die Garantie der Dienstqualität in einzelnen IP-Netzen zu kennen. Wichtige Punkte sind dabei die technischen Parameter für die Beschreibung der Dienstqualität, die Bedeutung der Fehlerrobustheit sowie die beiden grundsätzlichen Architekturen IntServ und DiffServ zur Gewährleistung der Dienstqualität innerhalb eines IP-Netzes.

Für die Beurteilung des neuen Netzdimensionierungsverfahrens in Kapitel 5 wird anhand eines Beispiels der Netzplanungsprozeß ausgeführt. Aus diesem Grund wurde der grundlegende Ablauf der Netzplanung eines IP-Netzes kurz erklärt.

Da die Topologie des realen Internets nicht bekannt ist, muß für die folgenden Untersuchung auf generierte Internet-Topologien zurückgegriffen werden. Als geeignet für das Erzeugen von Internet-Topologien wurde der GLP-Algorithmus ausgewählt und beschrieben, wie die in den folgenden Kapiteln verwendeten Internet-Topologien bestimmt werden.

Kapitel 3

Konvergenzverhalten von BGP

Das Interdomain-Routing mit BGP ist ein wichtiger Bestandteil des Internets und hat damit Einfluß auf die Gewährleistung der Interdomain-Dienstqualität. Vor allem die Dauer der BGP-Fehlerreaktion reduziert die Verfügbarkeit im Fall einer Interdomain-Routing-Änderung und beeinflußt damit die Wahrscheinlichkeit, mit der die Interdomain-Dienstqualität garantiert werden kann. Aus diesem Grund wird das Konvergenzverhalten von BGP in diesem Kapitel untersucht und dessen Eignung für die Sicherstellung der Interdomain-Dienstqualität bewertet.

Im ersten Abschnitt wird ein Überblick über die Randbedingungen des Interdomain-Routing im heutigen Internet gegeben. Nach der Feststellung der langen Dauer der BGP-Reaktion nach einer Topologieänderung werden die Gründe dafür erläutert. Daran schließt sich die Betrachtung des Standes der Wissenschaft bezüglich der Verbesserung des BGP-Konvergenzverhaltens an. Deren Bewertung ergibt, daß nur die Beschleunigung des Verteilens der BGP-Routing-Informationen uneingeschränkt zur Reduzierung der BGP-Konvergenzzeit führt. Die anschließende Abschätzung der unteren Grenze der Dauer der BGP-Reaktion auf Topologieänderungen ergibt ein Maß für die Bewertung der Beschleunigung des BGP-Konvergenzverhaltens. Um diese untere Grenze zu erreichen, werden die Verbesserungspotentiale des BGP-Zeitglieds MRAI und der BGP-Erweiterung EPIC untersucht. Dabei wird gezeigt, daß die Kombination von EPIC und das Deaktivieren des MRAI eine minimale Konvergenzzeit von BGP ermöglicht. Im letzten Abschnitt dieses Kapitels werden die gewonnenen Erkenntnisse für den Entwurf des erweiterten BGP-Routing-Protokolls BGP+ benutzt und die notwendigen Änderungen spezifiziert, um BGP+ ohne Kompatibilitätsprobleme im heutigen Internet einzuführen.

3.1 Interdomain-Routing im Internet

Um das Konvergenzverhalten von BGP einschätzen zu können, müssen die Randbedingungen, die an ein heutiges Interdomain-Routing-Protokoll im Internet gestellt werden, bekannt sein und werden deswegen im folgenden näher betrachtet. Die grundlegenden Eigenschaften des Internets lassen sich nach [BCS+00] wie folgt zusammenfassen:

1. Zusammenarbeit – Alle existierenden Netze, die IP nutzen, müssen untereinander Daten austauschen können.
2. Fehlerrobustheit – Die Kommunikation muß auch bei Ausfall einer Komponente sichergestellt werden.
3. Heterogenität – Alle IP-Netze müssen in die Internet-Architektur einbindbar sein, unabhängig von den eingesetzen Transporttechnologien.
4. Verteiltes Management – Die Ressourcen des Internets müssen verteilt verwaltet werden können.
5. Kosten – Die Internet-Architektur muß kosteneffizient sein.
6. Einfache Anbindung – Neue Netze müssen mit geringem Aufwand in die Internet-Architektur integriert werden können.
7. Abrechenbarkeit – Die im Internet verwendeten Ressourcen müssen zwischen den Netzen abrechenbar sein.

Neben diesen Anforderungen muß ein Interdomain-Routing-Protokoll im heutigen Internet auch dessen stetes Wachstum unterstützen. Dieses ist aus Sicht des australischen Netzbetreibers Telstra über die letzten siebzehn Jahre in Abbildung 3.1 dargestellt. Da ein Ende dieses Wachstums nicht abzusehen ist, muß ein Interdomain-Routing-Protokoll wie BGP auch in Zukunft mit den ständig zunehmenden Routing-Informationen umgehen können.

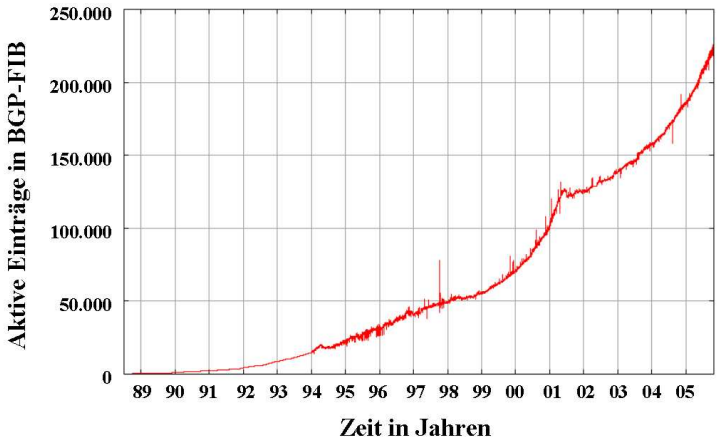


Abbildung 3.1: Wachstum des Internets aus Sicht des Betreibers Telstra [Tel05]

Der letzte Punkt in Abbildung 3.1 stammt vom 10. Oktober 2005. Zu diesem Zeitpunkt bestand das Internet, gesehen aus der Perspektive von Telstra, aus 20645 autonomen Systemen und 225787 aktiven Präfixen [Tel05], die jeder BGP-Router von Telstra zu verwalten hatte. Diese Werte verdeutlichen, daß nur ein Distance-Vector-Routing-Protokoll, das ohne die Kenntnis der kompletten Topologie auskommt, diese Skalierungsanforderung sicherstellen kann.

Neben der Skalierfähigkeit von BGP ist die individuelle Definition der BGP-Filterregeln, der sogenannten Policies (s. Abschnitt 2.1.2.3), eine weitere wichtige Eigenschaft des Interdomain-Routing im heutigen Internet und gleichzeitig ein Erfolgsfaktor von BGP. Soll BGP durch ein neues Interdomain-Routing-Protokoll ersetzen werden, muß dieses zukünftige Protokoll gleiche oder bessere Möglichkeiten bieten, um die Geschäftsinteressen der Betreiber abzubilden.

Der Einsatz der Policies hat verschiedene Konsequenzen auf das BGP-Routing. Aufgrund des Filterns und Veränderns der BGP-Routing-Informationen ist es möglich, daß bei der Bestimmung der besten Pfade von BGP längere Pfade kürzeren gegenüber bevorzugt werden. Außerdem erhält ein Betreiber eines autonomen Systems nur die BGP-Routing-Informationen, die die Policies der Betreiber entlang der durchlaufenen Pfade zulassen. Somit schränken die Policies die Verteilung der BGP-Routing-Informationen ein, und jedes autonome System entwickelt seine eigene Sicht auf die verfügbaren Pfade.

Da dies im Interesse der Geschäftsbeziehungen der einzelnen Betreiber geschieht, sind diese Einschränkungen als Vorteil zu werten. Allerdings kann jeder Betreiber die Policies für sein autonomes System ohne jede Koordination mit anderen autonomen Systemen definieren. Die Konsequenz dieses Mangels sind Instabilitäten beim BGP-Routing [GR01]. Als Instabilität oder Divergenz wird der Zustand von BGP bezeichnet, wenn BGP nicht einen eindeutigen Pfad zu einem Ziel findet und dann zwischen mehreren Pfaden hin- und herschaltet. Dieses Verhalten wird in der Literatur als *Route Flapping* bezeichnet.

Ein sehr anschauliches und einfaches Beispiel stellt das so genannte *Bad Gadget* [GSW02] dar. Hier zeigt T. Griffin anhand von vier autonomen Systemen, wie in Abhängigkeit der Reihenfolge der bevorzugten besten Pfade der einzelnen autonomen Systeme BGP ein stabile Lösung (*Good Gadget*) oder nicht (Bad Gadget) findet. An einer zweiten Beispielkonfiguration für fünf autonome Systeme veranschaulicht T. Griffin ein weiteres Problem. In diesem Fall kehrt BGP nach der Reparatur einer ausgefallenen Verbindung nicht zu dem Zustand vor dem Fehler zurück. Dies wird nur möglich, wenn alle betroffenen BGP-Verbindungen zurückgesetzt werden. Veröffentlicht hat T. Griffin dieses Beispiel unter dem Namen *Wedgies* [Gri04].

Mit diesen beiden Beispielen belegt T. Griffin, daß nur durch die Abstimmung der Policies der autonomen Systeme das Route Flapping verhindert werden kann. Dieselbe Schlußfolgerung wird in [MC05] gezogen, wo ein Überblick über die BGP-Divergenz gegeben wird.

Derzeit nimmt man das Risiko der BGP-Divergenz in Kauf. Aber will man die Dienstqualität auch zwischen autonomen Systemen gewährleisten, muß dafür eine Lösung gefunden werden. Eine Voraussetzung dafür ist die Offenlegung der Policies oder zumindest die Herausgabe formaler Daten, die das Abstimmen der Policies zwischen den Betreibern ermöglichen. Ein Lösungsvorschlag für so eine formale Beschreibung ist die *Routing Policy Specification Language* (RPSL) [AVG+99].

Das Wichtigste, das durch die Abstimmung der Policies sichergestellt werden muß, ist die *Valley-Free*-Eigenschaft der Pfade. Damit ist gemeint, daß die Einschränkungen der verallgemeinerten Policies (s. Abschnitt 2.1.2.4) in jedem Fall sichergestellt sind. So muß z.B. verhindert werden, daß ein Pfad von einem Betreiber-AS über sein Kunden-AS weiter über ein weiteres Betreiber-AS führt. Ist diese Anforderung nicht erfüllt, können zwei autonome Systeme über ein gemein-

sames Kunden-AS Verkehr austauschen, was im Gegensatz zu dem BGP-Routing sowie den Policies dieses Kunden-AS steht und es somit zu Widersprüchen in der Policies dieser drei autonomen Systemen kommen kann.

Daß die Bedingung der Valley-Free-Eigenschaft ausreicht, um das Route Flapping zu vermeiden, wird in [GR01] bewiesen und daraus ein allgemein gültiger Satz an einfachen Policies definiert, die dies sicherstellen (s. Abschnitt 2.1.2.4). Auf dieser Grundlage wird in [GS05] eine neue Algebra vorgestellt, die ein Interdomain-Routing-Protokoll ermöglicht, daß genau die Valley-Free-Eigenschaft der Pfade gewährleistet.

Wie in [MC05, JR05, GR01] ausgeführt wird, ist die Divergenz von BGP kein ausschließlich technisches Problem mehr, sondern die Lösung hängt vor allem von der Bereitschaft der Betreiber ab, ihre Policies zumindest teilweise offenzulegen bzw. die Freiheit bei der Definition ihrer Policies einzuschränken, damit die Policies aufeinander abgestimmt werden können. Aus diesem Grund wird das Thema in dieser Arbeit nicht weiter verfolgt, und es wird davon ausgegangen, daß nach einer Topologie-Änderung BGP immer einen stabilen Pfad zu den betroffenen Präfixen findet.

3.2 Zeitverhalten von BGP bei Topologieänderungen

BGP ist, wie jedes andere adaptive Routing-Protokoll auch, in der Lage, auf Änderungen zu reagieren. Beim Interdomain-Routing gibt es im wesentlichen drei Gründe [MWA02, WMR+05], warum sich ein Pfad zu einem Ziel ändert und BGP nach einem neuen besten Pfad suchen muß. Entweder ist eine Komponente ausgefallen, es ist ein Fehler in der Konfiguration von BGP aufgetreten oder der Interdomain-Verkehr muß an neue Gegebenheiten angepaßt werden.

Der Prozeß der Umlegung von Interdomain-Verkehr wegen geänderten Bedingungen wird auch als Interdomain-Verkehrsmanagement bezeichnet und wird ausführlich in [Uhl04] untersucht. Für die folgenden Untersuchungen ist es aber unerheblich, ob sich ein Pfad wegen des Interdomain-Verkehrsmanagements oder eines Fehlers ändert. In beiden Fällen muß BGP reagieren, wobei die Anforderungen an die Reaktionsgeschwindigkeit bei einem Ausfall deutlich höher sind. Deswegen werden im folgenden diese beiden Szenarien als Reaktion von BGP auf eine Topologieänderung zusammengefaßt.

Die Erkennung und Behebung einer falschen BGP-Konfiguration kann BGP nicht selbst durchführen, da das Fehlverhalten innerhalb von BGP begründet liegt. Zur Lösung der fehlerhaften Konfiguration ist das Eingreifen des Netzbetreibers unter Zuhilfenahme von zusätzlichen Werkzeugen notwendig, wie z.B. die Überwachung der Eingabe der Konfigurationsdaten. Wie mit diesem Fehlverhalten umzugehen ist, ist ein eigenes Problem mit ganz anderen Annahmen über das Verhalten von BGP und wird deswegen nicht weiter untersucht.

3.2.1 Dauer der BGP-Reaktion auf Topologieänderungen

Wie in Abschnitt 2.2 ausgeführt wurde, spielt die Dauer der BGP-Reaktion auf Topologieänderungen eine entscheidende Rolle bei der Gewährleistung der Dienstqualität, insbesonders bei der Sicherstellung der Verfügbarkeit. Deswegen wird im folgenden diese Dauer näher untersucht und als Konvergenzzeit bezeichnet.

Unter der Konvergenzzeit wird die Zeitdauer zwischen dem ersten Auftreten einer Topologieänderung und dem Wiederherstellen eines stabilen Zustandes verstanden. Damit beinhaltet sie die Erkennung der Topologieänderung, die Neuberechnung der betroffenen Pfade, die Benachrichtigung der Nachbarknoten und das Einrichten der neu gefundenen Pfade in die FIB des jeweiligen Routers [Por06].

In den folgenden Untersuchungen geht es hauptsächlich um die Konvergenz von BGP. Allerdings kann das Intradomain-Routing-Protokoll in Abhängigkeit der Routing-Änderung auf diese reagieren und somit wird ein Konvergenzprozeß innerhalb des betroffenen autonomen Systems initiiert.

Die Dauer der BGP-Konvergenz wurde erstmalig von Labovitz gemessen, indem von einem autonomen System aus die Bekanntgabe verschiedener Topologieänderungen im realen Internet angestoßen und die Reaktionszeiten gemessen wurden. Die ermittelten Ergebnisse wurden in [LAB+01] veröffentlicht. Dabei wurde die Konvergenz von BGP anhand der folgenden vier Szenarien untersucht:

1. der Ausfall eines Präfixes – $\mathrm{t_{down}}$
2. die Bekanntgabe eines unbekannten Präfixes – $\mathrm{t_{up}}$
3. der Verbindungsausfall und das Umschalten auf einen Backup Pfad – $\mathrm{t_{long}}$
4. ein neuer bester Pfad wird bekannt gegeben – $\mathrm{t_{short}}$

Die Fehlerszenarien $\mathrm{t_{up}}$ und $\mathrm{t_{short}}$ stellen jeweils das Komplementärereignis zu $\mathrm{t_{down}}$ bzw. $\mathrm{t_{long}}$ dar, da in diesen Fällen die Topologieänderung wieder aufgehoben wurde, was auch als die Reparatur eines Fehlers verstanden werden kann. Für diese vier Fehlerfälle sind die in [LAB+01] ermittelten Konvergenzzeiten von BGP im realen Internet in Abbildung 3.2 dargestellt.

Diese Meßergebnisse zeigen ganz deutlich die sehr lange BGP-Konvergenzzeit beim Auftreten eines Fehlers. Wenn ein Präfix nicht mehr erreichbar ist ($\mathrm{t_{down}}$), haben nach drei Minuten erst etwas mehr als 70% aller autonomen Systeme dieses Präfix aus ihren Routing-Tabellen gelöscht. Bei den verbliebenen autonomen Systemen hat die Reaktionsdauer bis zu 15 Minuten betragen. Existiert ein Backup Pfad ($\mathrm{t_{long}}$), gelingt das Umschalten auf diesen rund 20 Sekunden schneller verglichen mit dem Fall, wenn ein Präfix nicht mehr erreichbar ist.

Wird der Fehler behoben und ist der ursprüngliche beste Pfad wieder verfügbar, gelingt es BGP, diesen wesentlich schneller wieder zu benutzen. Im Durchschnitt dauert dies 90 Sekunden, wobei zwischen den beiden Fällen $\mathrm{t_{up}}$ und $\mathrm{t_{short}}$ kein Unterschied gemessen wurde.

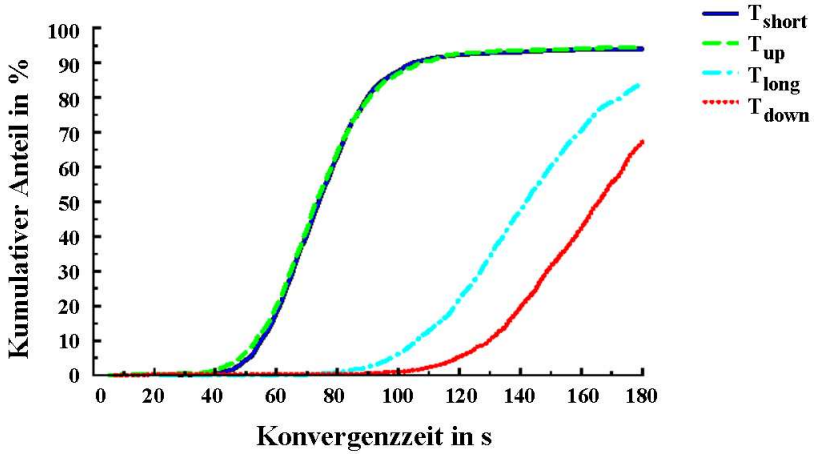


Abbildung 3.2: Meßwerte der BGP-Konvergenzzeit nach [LAB+01]

Bei der Arbeit, die in [MGV+02] beschrieben ist, wurden die Messungen von Labovitz zu einem späteren Zeitpunkt wiederholt, und dabei konnten die Ergebnisse aus [LAB+01] bestätigt werden.

Diese Konvergenzzeiten sind viel zu langsam und bedeuten für Dienste, die eine dedizierte Dienstqualität erfordern, den Abbruch der Datenübertragung und damit die Nichteinhaltung der zugesicherten Dienstqualität.

3.2.2 Ablauf der BGP-Konvergenz

Um die deutlichen Unterschiede in den von Labovitz gemessenen Konvergenzzeiten zu verstehen, ist ein detaillierter Blick auf den zeitlichen Ablauf der BGP-Reaktion auf Topologieänderungen notwendig.

3.2.2.1 Nichterreichbarkeit eines Präfixes

Wie BGP auf einen Ausfall reagiert, bei dem ein Präfix nicht mehr erreichbar ist, soll anhand des Beispiels in Abbildung 3.3 erklärt werden.

In diesem Beispiel fällt die Verbindung zwischen $\mathrm{AS_D}$ und $\mathrm{AS_1}$ aus, und $\mathrm{AS_D}$ ist dadurch für die restlichen autonomen Systeme nicht mehr zu erreichen. Der Ablauf der BGP-Fehlerreaktion soll aus Sicht des $\mathrm{AS_7}$ beschrieben werden, das die folgenden drei Pfade zu dem Ziel $\mathrm{AS_D}$ kennt – $< \mathrm{AS_D AS_1 AS_3} >$, $< \mathrm{AS_D AS_1 AS_2} >$, $< \mathrm{AS_D AS_1 AS_4 AS_6} >$.

Diese Notation beschreibt die durchlaufenen autonomen Systeme in der Reihenfolge, wie die BGP-UPDATE-Nachrichten den Pfad passieren. Aus diesen drei Pfaden wählt $\mathrm{AS_7}$ den für sich besten Pfad aus, wobei die Reihenfolge der Pfade die Rangfolge beschreibt.

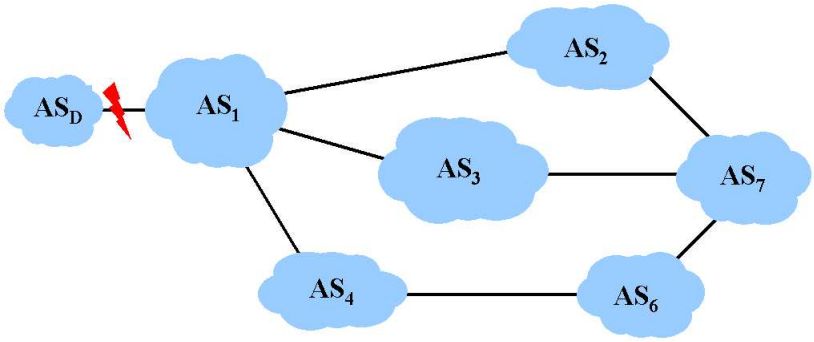


Abbildung 3.3: BGP-Pfadtesten im Fehlerfall

Nachdem AS_1 den Ausfall der Verbindung zu AS_D erkannt hat, sendet es an seine Nachbar-AS eine BGP-UPDATE-Nachricht, die diese Information enthält. Der weitere Verlauf hängt von der Reihenfolge ab, in der AS_7 von diesem Fehler erfährt, und im folgenden wird der schlechteste Fall angenommen.

Geht man davon aus, daß AS_7 zuerst von AS_3 eine BGP-UPDATE-Nachricht mit der Information erhält, daß AS_D nicht mehr erreichbar ist, führt dies zur Auswahl des zweitbesten Pfads $< \mathrm{AS}_\mathrm{D}\mathrm{AS}_1\mathrm{AS}_2 >$ als besten Pfad. Im Anschluß empfängt AS_7 von AS_2 eine BGP-UPDATE-Nachricht mit dem Inhalt, daß AS_2 AS_D nicht mehr erreichbar kann. Nachdem die zwei besten Pfade nicht mehr gültig sind, wählt AS_7 den dritten Pfad $< \mathrm{AS}_\mathrm{D}\mathrm{AS}_1\mathrm{AS}_4\mathrm{AS}_6 >$ als besten Pfad aus. Erst wenn AS_7 von AS_6 die BGP-UPDATE-Nachricht erhält, in der AS_6 mitteilt, daß es AS_D nicht mehr erreicht, erkennt AS_7, daß AS_D generell nicht mehr erreichbar ist.

Dieses Verhalten wird in der Literatur als Pfadtesten von BGP beschrieben, da BGP erst realisiert, daß das betroffene Präfix nicht mehr erreicht werden kann, nachdem alle existieren Pfade als ungültig erkannt worden sind.

Überträgt man das obige Beispiel auf eine größere Topologie, hat dies zur Folge, daß eine sehr große Anzahl an BGP-UPDATE-Nachrichten verschickt werden. Jedes autonome System leitet die Information über die Änderung eines besten Pfads sofort an seine Nachbar-AS weiter, sobald es die entsprechende Information erhalten hat. Je häufiger sich die Pfadinformationen ändern, um so mehr Routing-Informationen werden von BGP verteilt und somit entsteht eine riesige Welle an BGP-UPDATE-Nachrichten. Erst wenn alle ungültigen Pfade als solche erkannt wurden, endet die BGP-Reaktion. Aufgrund dieser Menge an BGP-UPDATE-Nachrichten, die die BGP-Router verarbeiten und weiterleiten müssen, kommt es zu der langen Konvergenzzeit von BGP. Damit ist das BGP-Pfadtesten auch die Ursache für die langsamen Meßergebnisse in [LAB+01], wenn ein Präfix nicht mehr erreicht werden kann.

3.2.2.2 Bekanntgabe eines unbekannten Präfixes

Beim Komplementärereignis zu t_down wird der Fehler, der zu der Nichterreichbarkeit des Präfixes geführt hat, repariert. Somit entspricht dieser Fall dem Bekanntgeben eines neuen, bisher

unbekannten Präfixes. Dabei informiert jedes autonome System seine Nachbar-AS über das neue Präfix, und jeder BGP-Router muß jeweils den für ihn besten Pfad auswählen und gegebenfalls weiterleiten.

Somit tritt das Phänomen des Pfadtestens hier nicht auf, und beim Verteilen der BGP-Routing-Informationen werden drastisch weniger BGP-UPDATE-Nachrichten generiert als beim Ausfall eines Präfixes. Aus diesem Grund fällt die Konvergenzzeit von BGP deutlich schneller aus, wie die Meßwerte von Labovitz bestätigen.

3.2.2.3 Änderung eines bekannten Pfads

Wenn sich in einem BGP-Router ein bester Pfad zu einem Ziel ändert und ein weiterer Pfad zur Auswahl steht, kommt es auf die Ursache der Änderung an, wie die Reaktion von BGP verläuft. Wird der bisher beste Pfad ungültig und der alternative Pfad stellt einen gültigen Ersatzpfad dar, handelt es sich um das Fehlerszenario $\mathrm{t_{long}}$. Hierbei muß BGP den bisherigen besten Pfad löschen und durch den Ersatzpfad ersetzen.

Ist der beste Pfad weiterhin gültig und der neue alternative Pfad besser im Sinne der Bestimmung des besten Pfads (s. Abschnitt 2.1.2.2), dann liegt das Fehlerszenario $\mathrm{t_{short}}$ vor. Die Ursachen für diesen Fall können z.B. das Verschieben von Verkehr in einem autonomen System sein, um dessen Anforderungen besser gerecht zu werden, oder ein Fehler, bei dem ein Ersatzpfad verfügbar war, wird behoben.

Da im Fall von $\mathrm{t_{long}}$ der bisher beste Pfad gelöscht werden muß, tritt auch das Problem des Pfadtestens auf. Allerdings nimmt das Pfadtesten nicht dieselben Ausmaße wie bei einem kompletten Ausfall eines Präfixes an. Da ein gültiger Ersatzpfad existiert, müssen nicht alle Pfade, die zu diesem Präfix führen, gelöscht werden. Nur die ungültigen Pfade, die besser als der Ersatzpfad sind, müssen als ungültig erkannt werden. Somit ist die Konvergenzzeit von BGP in diesem Fehlerszenario kürzer als bei der Nichterreichbarkeit eines Präfixes.

Muß BGP einen immer noch gültigen Pfad durch einen besseren ersetzen, ähnelt das Verhalten von BGP dem Fehlerszenario $\mathrm{t_{up}}$, denn hierbei wird kein Pfad ungültig, sondern ein besserer Pfad verdrängt einen anderen. Damit liegt die Konvergenzzeit auch in der gleichen Größenordnung wie im Fall von $\mathrm{t_{up}}$ und ist deutlich kürzer als die Konvergenzzeit von BGP, wenn Pfade als ungültig erkannt werden müssen. Allerdings kann die Konvergenzzeit für das Fehlerszenario $\mathrm{t_{short}}$ schwanken, da nicht zwangsläufig alle autonomen Systeme in die Reaktion von BGP involviert sein müssen. Stellt für ein autonomes System der neue Pfad keinen besseren Pfad als der aktuelle dar, dann gibt es die Informationen über diesen Pfad nicht an seine Nachbar-AS weiter, und diese sind von der Routing-Änderung nicht betroffen.

3.3 Verbesserung des BGP-Konvergenzverhaltens

Nach der Feststellung der langen Konvergenzzeit von BGP im letzten Abschnitt und der Erklärung der BGP-Reaktion in Abhängigkeit der vier Fehlerszenarien werden im folgenden die

wichtigsten existierenden Ansätze zur Verbesserung des Konvergenzverhaltens von BGP vorgestellt und am Ende bewertet.

3.3.1 Schnelle Fehlererkennung

Bevor ein Protokoll auf einen Fehler reagieren kann, muß dieser erst als solcher erkannt werden. Für die Fehlererkennung existieren verschiedene Ansätze, und die meisten Technologien beinhalten eigene Mechanismen. Da bei der Datenübertragung mehrere Technologien gemeinsam eingesetzt werden, wie z.B. IP über Ethernet über SDH, ist es sinnvoll, die Fehlererkennung dieser verschiedenen Technologien aufeinander abzustimmen. Wie diese Zusammenarbeit realisiert werden kann, wird ausführlich in [Aut02] Kapitel 3.9 untersucht.

Trotz der Fehlererkennungsmechanismen der Technologien, auf denen die Kommunikation mit BGP aufsetzt, ist es erforderlich, daß BGP seine eigene Fehlererkennung besitzt. Zum einem kann BGP andere Fehler erkennen als ein Mechanismus, der im Schichtenmodell unterhalb von BGP arbeitet, und zum zweiten bleibt BGP so unabhängig von der Fehlererkennung anderer Technologien.

Deswegen hat man auch für BGP in [RLH06] eine eigene Fehlererkennung, die auf dem Hello-Protokoll basiert, definiert. Dabei sendet jeder BGP-Router periodisch eine BGP-KEEPALIVE-Nachricht an seine Nachbarn, und wenn er innerhalb einer bestimmten Zeit keine Nachricht erhält, geht er davon aus, daß die entsprechende Verbindung nicht mehr funktioniert. Als Standardwert versendet BGP alle 30 Sekunden eine BGP-KEEPALIVE-Nachricht, und ein Fehler wird erkannt, wenn innerhalb des Zeitglieds *HoldTime* (Standardwert 90 Sekunden) keine BGP-KEEPALIVE-Nachricht oder eine BGP-UPDATE-Nachricht empfangen wird. Somit bestimmt bei BGP der Wert des Zeitglieds HoldTime die Dauer der Fehlererkennung.

Um diese Dauer deutlich zu verkürzen und nicht für jedes IP-Protokoll eine eigene Fehlererkennung zu definieren, wird bei der IETF das Protokoll *Bidirectional Forwarding Detection* (BFD) entwickelt [KW05]. Die BFD ist ein Mechanismus, der ebenfalls auf dem Hello-Protokoll basiert und sehr kleine Zeiten für das Senden der periodischen Nachrichten ermöglicht. Somit sollte mittels BFD eine Fehlererkennung in weniger als einer Sekunde möglich werden. In [ZDL05] wird auch an einer Schnittstelle für BGP gearbeitet, damit BGP ebenfalls von der schnellen Fehlererkennung der BFD profitieren kann. Somit sollte die BFD ein wichtiger Bestandteil von jedem Mechanismus sein, der zum Ziel hat, die Konvergenzzeit von BGP zu verringern.

3.3.2 Unterdrückung der Oszillation der Pfade

Wie in Abschnitt 3.1 erwähnt wurde, kann es während der BGP-Reaktion auf eine Topologieänderung zu sich ständig ändernden Pfaden zu einem Ziel kommen. Um dieses Problem zu beheben, wurde bei der IETF die BGP-Erweiterung *Route Flap Damping* (RFD) [VCG98] standardisiert.

Bei dieser Erweiterung wird jede Änderung eines Pfads zu einem Ziel mit einer Strafe p belegt. Welche Strafen für das Löschen eines Präfixes, für die Wiederbekanntgabe eines Präfixes und für das Ändern eines Attributs vergeben werden, ist in den ersten drei Spalten von Tabelle 3.1 angegeben. Übersteigt die Summe der Strafen eines Präfixes P eine definierte Schwelle (*Cutoff*), werden alle empfangenen Routing-Informationen für dieses Präfix unterdrückt und somit nicht mehr ausgewertet. Empfängt ein BGP-Router keine weiteren Routing-Informationen für das betroffene Präfix, reduziert das RFD nach der Gleichung 3.1 aus [MGV+02] die Strafe dieses Präfixes, wobei N den Nachbar des autonomen Systems beschreibt, H für die Halbwertzeit steht und λ sich aus der Gleichung 3.2 ableitet.

$$\mathrm{p}\,[\mathrm{P},\mathrm{N}]\,(\mathrm{t}') = \mathrm{p}\,[\mathrm{P},\mathrm{N}]\,(\mathrm{t}) * \mathrm{e}^{-\lambda(\mathrm{t}'-\mathrm{t})} \tag{3.1}$$

$$\mathrm{e}^{-\lambda \mathrm{H}} = 0,5 \tag{3.2}$$

Unterschreitet die Höhe der Strafe die Wiederverwendungsschwelle (s. Tabelle 3.1), werden neue Routing-Informationen dieses Präfixes wieder angenommen und verarbeitet. Somit ergibt sich ein zeitlicher Verlauf der Strafe für ein unterdrücktes Präfix, wie er in Abbildung 3.4 zu sehen ist.

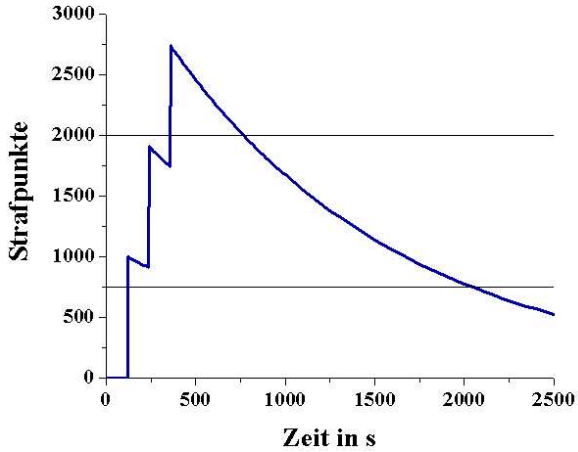


Abbildung 3.4: Zeitverhalten des RFD

Die Höhe der Strafen für die empfangenen Nachrichten, die Größe der beiden Schwellen sowie die maximale Unterdrückung eines Präfixes sind konfigurierbar, und damit läßt sich das Verhalten der RFD beeinflussen. Die von zwei Router-Herstellern empfohlenen Einstellungen sind in Tabelle 3.1 aufgeführt.

Allerdings sind diese Werte zu aggressiv gewählt, und die RFD unterdrückt auch gültige Routing-Informationen, wie in [BGM02, MGV+02] detailliert untersucht wurde. Die Meßergebnisse von [Mae03] bestätigen dieses Verhalten und zeigen zusätzlich, daß sich dadurch die Konvergenzzeit von BGP auf mehr als eine Stunde verlängern kann. Diese Dauer entspricht dem Standardwert für die maximale Zeit für die Unterdrückung eines Präfixes (s. Tabelle 3.1).

Hersteller	Präfix Löschen	Wiederbe-kanntgabe	Attribute ändern	Cutoff Schwelle	Halbwertzeit H in min	Wiederverwen-dungsschwelle	Max. Unterdrük-kungszeit in min
Cisco	1000	0	500	2000	15	750	60
Juniper	1000	1000	500	3000	15	750	60

Tabelle 3.1: Standardeinstellungen der RFD

Um das ungewollte Unterdrücken von gültigen Routing-Informationen zu vermeiden, wird in [MGV+02] das verbessertes *Selective Route Flap Damping* (SRFD) vorgestellt. Allerdings wird in [DCK+04] gezeigt, daß auch SRFD das Problem nicht löst, und deswegen wird ein weiterer Algorithmus, das RFD+, vorgeschlagen. Dessen Wirkungsweise läßt erwarten, daß RFD+ das Problem der sich ständig ändernden Pfade lösen kann. Diese Annahme wird in [DCK+04] mit simulativen Untersuchungen belegt, allerdings ist die Überprüfung der Richtigkeit an realen BGP-Daten ein noch offener Punkt. Trotzdem sollte RFD+ eingesetzt werden, um das Route Flapping zu verhindern.

3.3.3 Einführung von Redundanz

Eine weitere heute schon eingesetzte Maßnahme für die Verbesserung der BGP-Konvergenz ist die Erhöhung der Verfügbarkeit aller am BGP-Routing beteiligten Komponenten durch redundante Auslegung. Damit ändert man nicht das Konvergenzverhalten von BGP, sondern reduziert die Wahrscheinlichkeit, daß BGP auf einen Ausfall reagieren muß. Wie die dabei entstehende Kette von hoch verfügbaren Einzelteilen aufgebaut werden kann und welche Vorteile dies mit sich bringt, ist u.a. in [**ESL+05**] untersucht.

Diese hoch verfügbare Kette kann aber Fehler nicht verhindern und berücksichtigt Pfadänderungen aufgrund von Interdomain-Verkehrsmanagementmaßnahmen nicht. Genausowenig können die Auswirkungen von Intradomain-Routing-Entscheidungen auf das Routing mit BGP verhindert werden, die ausführlich in [TSG+04] untersucht wurden. Diese als *Hot-Potato* Routing bezeichneten Effekte mindern die Verfügbarkeit eines Pfads durch ein autonomes System, ohne daß ein Interdomain-Fehler im eigentliche Sinne auftritt.

3.3.4 Lokale Ersatzpfade

Die Vermeidung der BGP-Reaktion, bei der alle autonomen Systeme über die Topologieänderung informiert werden, stellt eine weitere interessante Idee dar, um die Konvergenzzeit von BGP drastisch zu verringern. Dafür muß im Fall einer Topologieänderung lokal ein Ersatzpfad bereitgestellt werden, wie anhand des Beispiels in Abbildung 3.5 erklärt werden soll. Dabei erkennt AS_4 den Ausfall der Verbindung $< \mathrm{AS}_4\mathrm{AS}_6 >$ und leitet den Verkehr über den lokalen Ersatzpfad $< \mathrm{AS}_4\mathrm{AS}_3\mathrm{AS}_6 >$ weiter.

Um den lokalen Ersatzpfad zu nutzen, ist keine Absprache notwendig, und AS_4 kann lokal auf den Ausfall reagieren, indem es den betroffenen Verkehr über den lokalen Ersatzpfad umleitet.

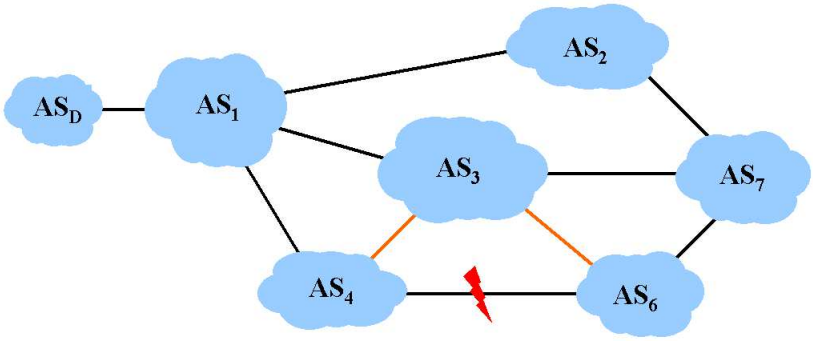


Abbildung 3.5: Fehlerreaktion mittels FaSRo

Wie in Abschnitt 2.1.2.2 beschrieben, nutzt BGP bei der Suche der besten Pfade das Attribut AS_PATH, das die von dem Pfad durchlaufenen autonomen Systeme enthält. Wird diese Eigenschaft wie im Fall der lokalen Ersatzpfade ohne Kenntnis der anderen autonomen Systeme verändert, kann es beim BGP-Routing zu Inkonsistenzen kommen. Deswegen müssen alle autonomen Systeme nach einer gewissen Zeit doch über die Topologieänderung informiert werden.

Um dennoch das Konzept der lokalen Ersatzpfade anzuwenden, wird beim *Fast Scoped Rerouting* (FaSRo) [BLS+03] ein Zeitglied eingeführt, das festlegt, wie lange der lokale Ersatzpfad genutzt werden kann. Als Größe für dieses Zeitglied werden zehn Minuten vorgeschlagen. Basierend auf den Aussagen von [MWA02] kann FaSRo damit auf rund 40% der Topologieänderungen lokal reagieren, da die Dauer dieser Fehlerfälle kleiner als zehn Minuten ist. In diesen Fällen ermöglicht FaSRo eine sehr schnelle Fehlerreaktion mit einer Dauer von 500ms. Allerdings wird in den restlichen 40% der Fehler die Standard-BGP-Fehlerreaktion initiiert, und dabei kommt zu der normalen BGP-Konvergenzzeit auch noch die Zeit für die Nutzung des lokalen Ersatzpfads hinzu.

Ein zu FaSRo sehr ähnlicher Ansatz wird in [BFF05] vorgestellt. Allerdings scheint es hier möglich zu sein, das Zurückfallen auf das Standard-Konvergenzverhalten von BGP zu verhindern, was aber nicht näher beschrieben ist und somit nicht nachvollzogen werden kann.

Die Idee der lokalen Ersatzpfade wird derzeit auch bei der IETF für das Intradomain-Routing unter dem Namen *IP Fast Reroute* aufgegriffen [Sha05a, AZ05]. Wie in [HCG06] gezeigt wird, hat auch IP Fast Reroute das Problem, daß nicht in jedem Fall alle Fehler betrachtet werden können. Allerdings werden hier Verfahren vorgeschlagen, die es ermöglichen, für alle Fehlerfälle lokale Ersatzpfade bereitzustellen. Diese Vorschläge lassen sich aber nicht auf das Interdomain-Routing übertragen, und damit verringert die Verwendung von lokalen Ersatzpfaden die BGP-Konvergenzzeit nur in einigen Fehlerfällen.

3.3.5 Vermeiden des BGP-Pfadtestens

In Abschnitt 3.2.2 wurde das Pfadtesten als eine der Hauptursachen der langen Konvergenzzeit von BGP eruiert. Aufgrund dessen werden im folgenden einige Ansätze, die dieses Problem

lösen wollen, untersucht.

Das *Ghost Flushing* [BBAS03] versucht durch aggressiveres Löschen der ungültigen Pfade die Konvergenzzeit von BGP zu verkürzen. Dazu werden die BGP-UPDATE-Nachrichten in zwei Etappen versendet. Als erstes werden die zu löschenden Präfixe mittels der BGP-UPDATE-Nachrichten, die nur aus dem Withdrawn-Teil bestehen, an alle autonomen Systeme verteilt. Erst im Anschluß daran werden die Ersatzpfade mittels der BGP-UPDATE-Nachrichten, die nur die geänderten Pfade der betreffenden Präfixe enthalten, bekanntgegeben. Aber dieser Ablauf kann das Pfadtesten von BGP nicht verhindern und bietet nur eine geringfügige Verbesserung der BGP-Konvergenzzeit.

Eine Lösung, um das Pfadtesten von BGP zu vermeiden, ist die Mitgabe der Information über den Ursprung des Fehlers mit jeder BGP-UPDATE-Nachricht. In diesem Fall kann jeder betroffene BGP-Router anhand der in der BGP-UPDATE-Nachricht enthaltenen Information alle ungültigen Pfade sofort nach dem Erhalt der ersten Nachricht löschen. Diese Idee wurde in drei ähnlichen Ansätzen in [MVW+00, PAM+05, ZAL03] verwirklicht.

Allerdings weisen diese Lösungsvorschläge ein Fehlverhalten auf, das durch die Nichtbeachtung der internen Topologie eines autonomen Systems hervorgerufen wird, was anhand eines Beispiels kurz erklärt werden soll. Betrachtet man die Topologie in Abbildung 3.6, wo die Interdomain-Verbindung $< \mathrm{AS_D AS_1} >$ (brauner Pfeil) ausfällt, dann erkennen alle benachbarten autonomen Systeme $\mathrm{AS_1}$ als Ursprung des Fehlers. Basierend auf der Verteilung dieser Information können alle autonomen Systeme alle Pfade, die die Verbindung $< \mathrm{AS_D AS_1} >$ enthalten, löschen, was dem geplanten Verhalten entspricht.

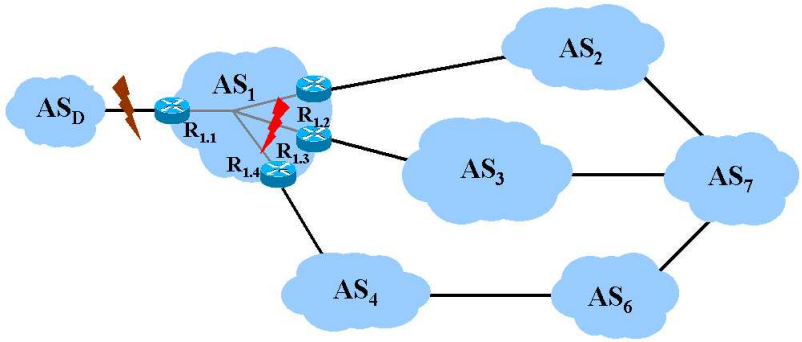


Abbildung 3.6: Ausfall einer Intradomain- (rot) und Interdomain- (braun) Verbindung

Allerdings ändert sich die Situation, wenn eine Verbindung innerhalb von $\mathrm{AS_1}$ ausfällt. Im Fall eines Fehlers der Verbindung zwischen $\mathrm{R_{1,1}}$ und $\mathrm{R_{1,3}}$ von $\mathrm{AS_1}$ (roter Pfeil) wird von den benachbarten autonomen Systemen ebenfalls $\mathrm{AS_1}$ als Fehlerursache erkannt und an alle anderen autonomen Systeme verteilt. Somit können die anderen autonomen Systeme zwischen dem internen und externen Fehlerfall nicht unterscheiden, da in beiden Fällen derselbe Fehlerursprung übermittelt wird. Damit sind diese autonomen Systeme im Fall des internen Fehlers nicht in der Lage, die beiden verfügbaren Ersatzpfade $< \mathrm{AS_D AS_1 AS_2 AS_7} >$ sowie $< \mathrm{AS_D AS_1 AS_4 AS_6 AS_7} >$ zu nutzen.

Dieses Problem wurde zuerst in [CDK+05] erkannt und behoben. Dazu wurde das *Enhanced Path Vector Protocol* (EPIC) [CDK+05] entwickelt, in dem die Pfadbeschreibung des Pfadvektors von BGP erweitert wird. Bei BGP wird ein Pfad zu einem Ziel als eine Folge von AS-Nummern beschrieben. Diese wird bei EPIC durch eine Liste von Nummern der Eingangs- und Ausgangs-Router der durchlaufenen autonomen Systeme ersetzt. Diese Nummern werden als *Forward Edge Sequence Number* (FESN) bezeichnet und sind innerhalb eines autonomen Systems eindeutig. Mit dieser Beschreibung und der Übermittlung der Fehlerursache ist es möglich, Fehler innerhalb eines autonomen Systems und zwischen verschiedenen autonomen Systemen zuverlässig zu detektieren und somit das Pfadtesten von BGP zu vermeiden.

Ein Nachteil von EPIC, der in [YMBB05] aufgezeigt wird, ist eine Einschränkung der Aggregation von Pfaden aufgrund der detaillierten Pfadbeschreibung mit der Liste der FESN. Man kann aber davon ausgehen, daß der Vorteil des Vermeidens des Pfadtestens von BGP überwiegt gegenüber der verschlechterten Aggregation.

3.3.6 Overlay-Netze

Einen Ansatz, der ebenfalls zur Reduzierung der BGP-Konvergenzzeit genutzt werden kann, stellt das Konzept der Overlay-Netze dar. Hierbei wird ein Netz aufgebaut, das auf einem bestehenden Netz aufsetzt und dessen Funktionalität nutzt. Der Vorteil der Overlay-Netze besteht darin, neue Funktionalitäten einzuführen, die unabhängig von dem darunterliegenden Netz sind und somit die Einschränkungen, wie z.B. die Kompatibilität, nicht beachtet werden müssen. Ein bekanntes Beispiel für ein Overlay-Netz sind die Peer-to-Peer-Netze, die das heutige Internet als Basis nutzen.

Das Konzept der Overlay-Netze läßt sich auch auf das BGP-Routing anwenden. Wie in [ACK03] gezeigt wird, kann man die Fehlerreaktion von BGP beschleunigen, indem das Overlay-Netz benutzt wird, um einen Ersatzpfad im voraus zu berechnen und im Fehlerfall auf diesen umzuschalten. Desweiteren werden in dem Overlay-Netz aus [ACK03] *Policy Agents* und *Policy Databases* eingerichtet, um eine Abstimmung der Policies zu ermöglichen und somit eine Divergenz aufgrund widersprechender Policies zu vermeiden.

Die Performanz eines Overlay-Netzes, wo von Anfang an mehrere Pfade zu einem Ziel berechnet werden, wird in [APS+04] untersucht. Dabei werden mehrere Varianten, bei denen unterschiedlich viele Pfade zu einem Ziel existieren, betrachtet und eine deutlich bessere Datenübertragung hinsichtlich der Verzögerung, des TCP-Durchsatzes und der Verfügbarkeit festgestellt.

Ebenfalls eine interessante Einsatzmöglichkeit der Overlay-Netze stellt die Gewährleistung der Interdomain-Dienstqualität dar, was in Abschnitt 4.2.6 noch näher untersucht und bewertet wird.

3.3.7 Bewertung der existierenden Lösungsvorschläge

Die in den letzten Abschnitten vorgestellten Mechanismen repräsentieren den Stand der Wissenschaft zur Verbesserung des Konvergenzverhaltens von BGP, wobei nicht alle die Reduzierung

der Konvergenzzeit als Ziel besitzen. So versucht das RFD, die negativen Einflüsse sich ständig wechselnder Pfade zu einem Ziel zu minimieren. Allerdings wird dies erst mit dem erweiterten RFD+ möglich. Auch die redundante Auslegung ändert nichts am eigentlichen Konvergenzverhalten von BGP, sondern reduziert die Auftrittshäufigkeit von Fehlern. Trotzdem stellen diese beiden Mechanismen wichtige Erweiterungen von BGP dar, um die Verfügbarkeit der BGP-Pfade zu erhöhen.

Um die Konvergenzzeit zu verkürzen, ist auf alle Fälle eine schnelle Fehlererkennung mittels der BFD notwendig. Aber für die weitere Reduzierung der Konvergenzzeit werden zusätzliche Ansätze benötigt. Die Nutzung eines lokalen Ersatzpfads ermöglicht eine sehr schnelle BGP-Reaktion. Wie in Abschnitt 3.3.4 erklärt wurde, ist dies aber nur in Fehlerfällen mit einer kurzen Dauer möglich.

Innerhalb eines autonomen Systems werden kurze Konvergenzzeiten erreicht, indem die Ersatzpfade vorab berechnet werden und somit beim Auftreten des Fehlers der Verkehr nur auf die Ersatzpfade umgelegt werden muß [RMD05]. Allerdings läßt sich dieser Ansatz auf das Interdomain-Routing nicht anwenden. Nicht nur, daß das Vorhalten von mehreren Pfaden zu einem Präfix die BGP-Routing-Tabellen um ein Mehrfaches vergrößert, ebensowenig kann das BGP-Routing sicherstellen, daß ein redundanter Ersatzpfad bei der Topologieänderung unverändert bleibt.

Anhand eines Beispiels, das in Abbildung 3.7 dargestellt ist, kann gezeigt werden, daß ein redundanter Ersatzpfad bei der entsprechenden Topologieänderung nicht derselbe bleibt.

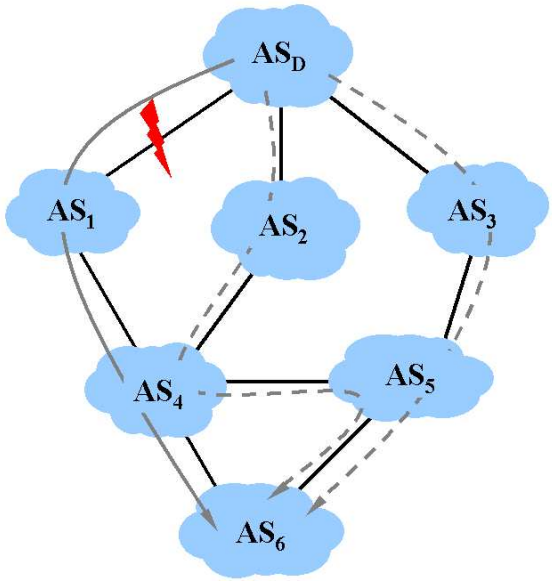


Abbildung 3.7: Änderungen eines BGP-Ersatzpfads nach einer Topologieänderung

Abbildung 3.7 zeigt ein Netz mit sieben autonomen Systemen, wobei von Interesse ist, über welchen Pfad AS_6 AS_D erreicht. AS_6 erhält via BGP von seinen Nachbar-AS deren besten

Pfad zu $\mathrm{AS_D}$. $\mathrm{AS_4}$ bevorzugt den Pfad $< \mathrm{AS_D AS_1 AS_4} >$, und der beste Pfad von $\mathrm{AS_5}$ lautet $< \mathrm{AS_D AS_3 AS_5} >$. Da diese beiden Pfade nur das Ziel-AS gemeinsam haben, sind es redundante Pfade aus Sicht von $\mathrm{AS_6}$. Damit sollte es möglich sein, im Fall eines Fehlers von einem der beiden Pfade auf den anderen Pfad umzuschalten.

Fällt nun die Verbindung zwischen $\mathrm{AS_D}$ und $\mathrm{AS_1}$ aus, kann $\mathrm{AS_6}$, sobald es davon erfährt, auf den Pfad $< \mathrm{AS_D AS_3 AS_5 AS_6} >$ umschalten, und wäre von dem Fehler nicht mehr betroffen. Aber aufgrund dieses Fehlers wählt $\mathrm{AS_4}$ den Pfad $< \mathrm{AS_D AS_2 AS_4} >$ zu seinem neuen besten Pfad und teilt dies $\mathrm{AS_6}$ und $\mathrm{AS_5}$ mit. Entscheidet nun $\mathrm{AS_5}$ diesen neuen Pfad als seinen neuen besten Pfad auszuwählen, ändert sich der redundante Ersatzpfad von $\mathrm{AS_6}$.

Für die Entscheidung von $\mathrm{AS_5}$ kann es viele verschiedene Gründe geben. Zum einem wählt BGP nicht zwangsläufig den kürzesten Pfad als besten Pfad aus. Zum anderen kann es aber auch sein, daß $\mathrm{AS_5}$ im allgemeinen die Pfade über $\mathrm{AS_4}$ bevorzugt, aber erst nach dem Fehler erfahren hat, daß $\mathrm{AS_4}$ $\mathrm{AS_D}$ erreichen kann. So kann z.B. $\mathrm{AS_1}$ verhindern, daß der Pfad über ihn nur an bestimmte autonome Systeme weitergegeben wird. Genauso gut können die Policies von $\mathrm{AS_5}$ bewirken, daß ein Pfad über $\mathrm{AS_1}$ von der Suche der besten Pfade ausgeschlossen wird.

Dieses Verhalten wird möglich, da alle Betreiber eines autonomen Systems ihre eigenen Routing-Entscheidungen anhand selbst definierter Präferenzen treffen, und das BGP-Routing nicht vorhersagbar ist. Das ganze wird noch erschwert, da ein autonomes System maximal nur die besten Pfade der anderen autonomen Systeme kennt und somit nur ein Teilwissen über die möglichen Pfade besitzt. Damit kann mittels vorausberechneter Ersatzpfade das BGP-Konvergenzverhalten von BGP nur bedingt reduziert werden.

So führen auch die Lösungsansätze der Overlay-Netze aus Abschnitt 3.3.6 nicht zum Erfolg, da auch sie nicht verhindern können, daß ein Ersatzpfad während einer Topologieänderung verändert wird. Natürlich könnte man mit dem Konzept der Overlay-Netze ein neues Routing-Protokoll aufbauen, das dieses Problem nicht aufweist. Aber das läßt sich nur realisieren, wenn man die Routing-Entscheidung der anderen autonomen Systeme vorhersagen kann. Da diese Forderung das BGP-Routing grundlegend verändert und man das komplette Routing, ohne auf Informationen von BGP zurückzugreifen, in allen Overlay-Knoten implementieren muß, kann man auch gleich das bestehende BGP ersetzen, was einfacher und effizienter sein dürfte.

Desweiteren ist der Abgleich der Policies mittels eines Overlay-Netzes ein guter und interessanter Ansatz. Allerdings kann ein Overlay-Netz nicht erzwingen, daß die Betreiber der autonomen Systeme ihre Policies offenlegen. Somit weist dieser Ansatz dieselben Nachteile auf wie die Konzepte, die direkt auf BGP aufbauen.

Als Konsequenz aus diesen Überlegungen ergibt sich nur eine Lösungsmöglichkeit, um die Konvergenzzeit von BGP zu reduzieren. Man muß den zeitlichen Ablauf, wie BGP die Routing-Informationen an alle autonomen Systeme verteilt, beschleunigen. Dabei kann der Einsatz von EPIC, womit das Pfadtesten von BGP vermieden wird, einen wichtigen Beitrag leisten. Die Untersuchungen in [CDK$^+$05] zeigen eine reduzierte Konvergenzzeit bei der Nichterreichbarkeit eines Präfixes und beim Umschalten auf einen Ersatzpfad bei vollvermaschten Topologien. Allerdings ist damit noch nicht geklärt, inwieweit EPIC eine allgemeingültige Verringerung der Konvergenzzeit von BGP im realen Internet zuläßt und wo die Grenzen bei der Reduzierung der

Konvergenzzeit liegen. Diesen Fragestellungen soll in den folgenden Abschnitten nachgegangen werden.

3.4 Untere Grenze der BGP-Konvergenzzeit

Nach der Feststellung, daß nur die Beschleunigung des Ablaufes der BGP-Konvergenz uneingeschränkt zu einer deutlichen Reduzierung der Konvergenzzeit führt, wird im folgenden eine Abschätzung der unteren Grenze der BGP-Konvergenzzeit vorgenommen, um ein Maß für die Bewertung der Verringerung der BGP-Konvergenzzeit zu erhalten.

Die Abschätzung der unteren und oberen Grenze in [LAB$^+$00] hilft nur bedingt weiter. Bei dieser Untersuchung wurde immer von einer vollvermaschten Topologie ausgegangen. Durch diese Annahme läßt sich zusätzlich die Anzahl der versendeten Nachrichten abschätzen, aber diese Werte lassen sich nicht auf das BGP-Routing im Internet übertragen, da die reale Internet-Topologie nicht vollvermascht ist (s. Abschnitt 2.4).

Wird der Ausfall einer Interdomain-Verbindung betrachtet, liefert diese Modellierung zu hohe Werte für die Konvergenzzeit und die Nachrichtenanzahl, da die Anzahl der vom Pfadtesten betroffenen Pfade wesentlich größer ist. Beim komplementären Ereignis sind die Werte wiederum zu gut, da der kürzeste Pfad immer eine Länge von einem AS-Hop hat. Im realen Internet liegt diese Länge zwischen drei und vier AS-Hops. Deswegen wird im folgenden ein neuer Ansatz entworfen, der auf der realen Internet-Topologie basiert.

Ausgangspunkt dieser Modellierung ist die Ausbreitung einer BGP-UPDATE-Nachricht entlang des entsprechenden Pfads. Je nach Wahl dieses Pfads repräsentiert dessen Ausbreitungsdauer die Konvergenzzeit von BGP, wobei das betrachtete Fehlerszenario den auszuwählenden Pfad vorgibt.

Betrachtet man das Szenario t_{down}, wo ein Präfix nicht mehr erreicht werden kann, muß der längste mögliche Pfad aller autonomen Systeme für die Abschätzung der BGP-Konvergenzzeit herangezogen werden. Allerdings ist die Bestimmung des längsten Pfads eines Graphen ein NP-hartes Problem. Existiert ein Ersatzpfad für eine ausgefallene Interdomain-Verbindung (Szenario t_{long}), bestimmt der längste Pfad, der aufgrund des Fehlers nicht mehr gültig ist, die BGP-Konvergenzzeit. Bei der Bekanntgabe eines neuen Präfixes oder eines besseren Pfads (Szenarien t_{up} bzw. t_{short}) ist der beste Pfad der Entscheidende. Um die BGP-Konvergenzzeit zu ermitteln, muß die Ausbreitung des besten Pfads mit der größten Länge berücksichtigt werden.

Unabhängig, welcher Pfad für die Abschätzung der Konvergenzzeit von BGP herangezogen wird, wird die Ausbreitungszeit einer BGP-UPDATE-Nachricht von zwei unterschiedlichen Zeiten beeinflußt, von der Sendedauer für das Versenden einer BGP-UPDATE-Nachricht zwischen zwei autonomen Systemen und von der Durchlaufzeit einer BGP-UPDATE-Nachricht durch ein autonomes System. Wie sich die Durchlaufzeit eines autonomen Systems zusammensetzt, wird im folgenden Abschnitt erläutert.

3.4.1 Durchlaufzeit

Als Durchlaufzeit (DLZ) wird in dieser Arbeit die Zeit bezeichnet, die eine BGP-UPDATE-Nachricht benötigt, um ein autonomes System zu passieren. Im einfachsten Fall besteht die DLZ aus drei verschiedenen Teilen, der Verweildauer im Eingangs-Router des autonomen Systems, der Übertragungszeit innerhalb des autonomen Systems und der Verweildauer im Ausgangs-Router. Die Verweildauer in einem BGP-Router faßt die Zeiten für die Berechnung des besten Pfads anhand der Routing-Informationen der BGP-UPDATE-Nachricht, für die Wartezeit in den Warteschlangen und weitere Verzögerungen, die beim Durchleiten einer BGP-Nachricht durch einen BGP-Router entstehen, zusammen.

Dem in Abbildung 3.8 veranschaulichten zeitlichen Ablauf des Passierens einer BGP-UPDATE-Nachricht durch ein autonomes System liegt die Annahme zugrunde, daß die betrachtete BGP-UPDATE-Nachricht Routing-Informationen enthält, die zu einem neuen besten Pfad führen und an die entsprechenden Nachbarn weitergegeben wird. Desweiteren wird davon ausgegangen, daß es sich um eine reguläre Topologieänderung handelt, wo das RFD nicht eingreift (s. Abschnitt 3.3.2).

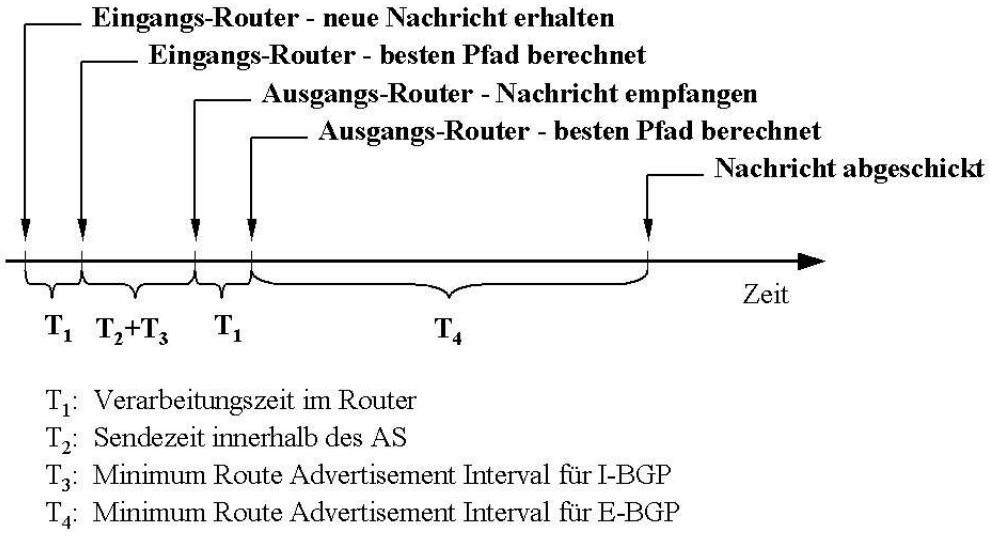


Abbildung 3.8: Durchlaufzeit einer BGP-UPDATE-Nachricht durch ein autonomes System

Die Zeit T_1 beschreibt die Verweildauer einer BGP-UPDATE-Nachricht in einem BGP-Router, und die Sendezeit T_2 besteht nicht nur aus der Übertragungszeit zwischen dem Eingangs-Router und dem Ausgangs-Router, sondern auch aus den Zeiten, die notwendig sind, um andere Router, die kein BGP unterstützen, zu passieren.

Als T_3 und T_4 wird das Zeitglied MRAI bezeichnet, das definiert wurde, um mehrere kurz hintereinander eintreffende BGP-UPDATE-Nachrichten gleichzeitig bei der Suche der besten Pfade zu berücksichtigen. Dies führt dazu, daß innerhalb einer MRAI-Periode nur eine BGP-UPDATE-Nachricht mit dem resultierenden besten Pfad pro Nachbar-Router verschickt wird, womit sich unter Umständen die Anzahl der versendeten BGP-UPDATE-Nachrichten verringert.

In [RLH06] sind zwei Anwendungsbereiche für den MRAI definiert, zum einem wird er innerhalb eines autonomen Systems, also im Eingangs-Router eines autonomen Systems, und zum zweiten beim Verlassen des autonomen Systems im Ausgangs-Router angewendet. Desweiteren werden unterschiedliche Standardwerte für diese beiden Zeitglieder vorgeschlagen. Basierend auf diesen Betrachtungen läßt sich die DLZ mit der Gleichung 3.3 beschreiben.

$$\mathrm{DLZ} = \mathrm{T}_1 + \mathrm{T}_2 + \mathrm{T}_3 + \mathrm{T}_1 + \mathrm{T}_4 \tag{3.3}$$

3.4.2 Berücksichtigung von Routen-Reflektoren

Bei der Bestimmung der DLZ durch die Gleichung 3.3 wurde von einer einfachen internen Struktur eines autonomen Systems ausgegangen. Besteht die interne Topologie eines autonomen Systems aus vielen BGP-Routern, kann es vorteilhaft sein, Routen-Reflektoren einzusetzen, wie es in Abschnitt 2.1.2.5 erläutert wurde. In diesem Fall wird eine BGP-UPDATE-Nachricht nicht direkt vom Eingangs-Router zum Ausgangs-Router geschickt. Sondern der Routen-Reflektor übernimmt das Weiterleiten der BGP-UPDATE-Nachrichten zwischen den entsprechenden BGP-Routern. Allerdings beschränkt sich die Funktionalität des Routen-Reflektors nicht nur auf das Weiterleiten von BGP-UPDATE-Nachrichten, sondern der Routen-Reflektor wertet die empfangenen BGP-UPDATE-Nachrichten auch aus, berechnet den besten Pfad, und falls er einen besseren Pfad gefunden hat, sendet er eine BGP-UPDATE-Nachricht an die entsprechenden BGP-Router. Somit muß bei der Berechnung der DLZ die Verweildauer einer BGP-UPDATE-Nachricht im Routen-Reflektor und der MRAI für I-BGP am Routen-Reflektor mit einkalkuliert werden, wie es in Abbildung 3.9 zu sehen ist.

Aufgrund des Routen-Reflektors wird die Sendezeit T_2 zwischen dem Eingangs- und Ausgangs-Router in zwei Teile geteilt, der Sendezeit $\mathrm{T}_{2\mathrm{a}}$ zwischen dem Eingangs-Router und dem Routen-Reflektor sowie der Sendezeit $\mathrm{T}_{2\mathrm{b}}$ zwischen dem Routen-Reflektor und dem Ausgangs-Router. Außerdem kommt noch eine weitere Verarbeitungszeit T_5 und ein Zeitintervall des MRAI für die interne BGP-Kommunikation hinzu, da der Routen-Reflektor ebenfalls die besten Pfade ermittelt. Wendet man diesen Einfluß eines Routen-Reflektors auf die Berechnung der DLZ aus Gleichung 3.3 an, erhält man die Gleichung 3.4.

$$\mathrm{DLZ} = \mathrm{T}_1 + \mathrm{T}_{2\mathrm{a}} + \mathrm{T}_3 + \mathrm{T}_5 + \mathrm{T}_3 + \mathrm{T}_{2\mathrm{b}} + \mathrm{T}_3 + \mathrm{T}_1 + \mathrm{T}_4 \tag{3.4}$$

Bei sehr großen Netzen besteht die Möglichkeit, auch mehrere Routen-Reflektoren einzusetzen (s. Abschnitt 2.1.2.5). Je nach Konfiguration können dann noch weitere Zeiten zu der DLZ hinzukommen. Werden mehrere Routen-Reflektoren in einer Hierarchieebene eingesetzt, dann muß zu Gleichung 3.4 eine weitere Verarbeitungszeit T_5, ein weiterer MRAI für I-BGP T_3 und die Sendezeit T_6 zwischen zwei Routen-Reflektoren hinzuaddiert werden. Dies führt zu der Gleichung 3.5.

$$\mathrm{DLZ} = \mathrm{T}_1 + \mathrm{T}_{2\mathrm{a}} + \mathrm{T}_3 + \mathrm{T}_5 + \mathrm{T}_3 + \mathrm{T}_6 + \mathrm{T}_5 + \mathrm{T}_3 + \mathrm{T}_{2\mathrm{b}} + \mathrm{T}_3 + \mathrm{T}_1 + \mathrm{T}_4 \tag{3.5}$$

Erfordert die interne Topologie eines autonomen Systems mehrere Hierarchieebenen an Routen-Reflektoren, müssen pro Routen-Reflektor die Verweildauer des entsprechenden Routen-Reflektors, ein MRAI für I-BGP und die zusätzliche Sendezeit zwischen den Routen-Reflektoren mit

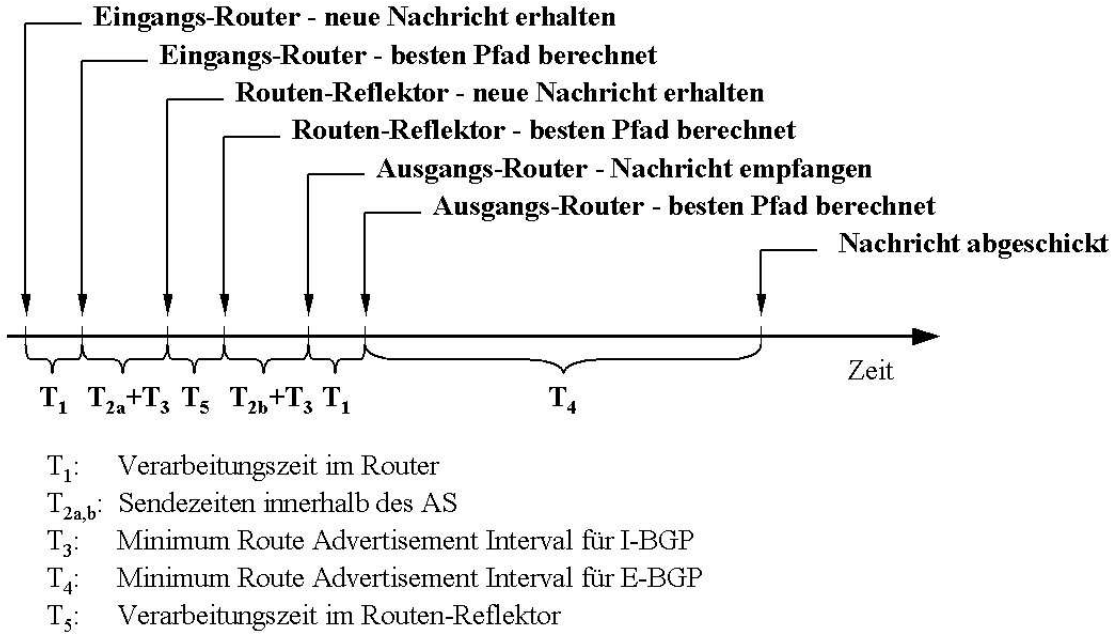


Abbildung 3.9: Modell der Durchlaufzeit eines autonomen Systems mit Routen-Reflektor

berücksichtigt werden. Der Einfachheit wegen wird im folgenden nur der einfache Aufbau mit einem Routen-Reflektor betrachtet.

3.4.3 Abschätzung der Durchlaufzeit

Nachdem in den letzten beiden Abschnitten die Zusammensetzung der DLZ in Abhängigkeit von der Benutzung von Routen-Reflektoren ermittelt wurden, soll nun die DLZ eines autonomen Systems abgeschätzt werden. Dazu müssen für die einzelnen Bestandteile reale Zeiten eingesetzt werden.

Die Dauer T_1, die eine BGP-UPDATE-Nachricht benötigt, um einen BGP-Router zu passieren, hängt stark vom betrachteten Router-Modell und der aktuellen Lastsituation ab. In [FKM+04] sind Messungen an einem Router des Herstellers Cisco vom Typ 12008 vorgenommen worden. Dabei wurde der Mittelwert für $\mathrm{T}_1 = 150\mathrm{ms}$ ermittelt. Aufgrund des Mangels anderer öffentlich verfügbarer Ergebnisse wird dieser Wert in den nachfolgenden Abschätzungen verwendet.

Für die beiden MRAI-Zeitglieder wurden die Standardwerte von 30s (E-BGP) bzw. 5s (I-BGP), wie in [RLH06] empfohlen, ausgewählt. Die Sendezeit innerhalb eines autonomen Systems wurde mit 30ms angenommen. Setzt man diese Werte in die Gleichung 3.3 ein, dann ergibt sich eine DLZ von 35,3s, wenn keine Routen-Reflektoren eingesetzt werden.

$$\mathrm{DLZ} = 150\mathrm{ms} + 30\mathrm{ms} + 5\mathrm{s} + 150\mathrm{ms} + 30\mathrm{s} = 35,3\mathrm{s}$$

Berücksichtigt man zusätzlich einen Routen-Reflektor, vergrößert sich nach Gleichung 3.4 die DLZ auf 40,5s.

$$\mathrm{DLZ} = 150\mathrm{ms} + 15\mathrm{ms} + 5\mathrm{s} + 150\mathrm{ms} + 15\mathrm{ms} + 5\mathrm{s} + 150\mathrm{ms} + 30\mathrm{s} = 40,5\mathrm{s}$$

Sollen noch weitere Routen-Reflektoren verwendet werden, sind für jeden weiteren Routen-Reflektor 5,15s und die zusätzliche Sendezeit hinzuzuaddieren.

3.4.4 Abschätzung der BGP-Konvergenzzeit

Basierend auf der DLZ und in Abhängigkeit der Pfadlänge kann man die Ausbreitungsdauer einer BGP-UPDATE-Nachricht entlang eines bestimmten Pfads abschätzen. Welcher Pfad für welchen Fehlerfall ausgewählt werden muß, wurde schon am Anfang von Abschnitt 3.4 erläutert.

In Abhängigkeit der Pfadlänge hops der zu betrachtenden Pfade läßt sich die Konvergenzzeit KZ abschätzen, indem man diese aus der Zeit für die Fehlererkennung (FE), der DLZ der durchlaufenen autonomen Systeme und der Sendezeiten zwischen diesen autonomen Systemen ($\mathrm{T_V}$) zusammensetzt, wie es in Gleichung 3.7 beschrieben ist.

$$\mathrm{KZ} = \mathrm{FE} + \mathrm{hops} * \mathrm{DLZ} + (\mathrm{hops} - 1) * \mathrm{T_V} + \mathrm{T_V} + \mathrm{T_1} \tag{3.6}$$

$$\mathrm{KZ} = \mathrm{FE} + \mathrm{hops} * (\mathrm{DLZ} + \mathrm{T_V}) + \mathrm{T_1} \tag{3.7}$$

Der letzte Teil der Gleichung 3.6 berücksichtigt, daß das letzte autonome System nach dem Erhalt der relevanten BGP-UPDATE-Nachricht die daraus resultierende Routing-Information noch an die Nachbarn-AS verteilt. Dies führt zu der zusätzlichen Addition der Interdomain-Sendezeit $\mathrm{T_V}$ und der Verweildauer der BGP-UPDATE-Nachricht im Eingangs-Router der Nachbar-AS ($\mathrm{T_1}$ aus Gleichung 3.3).

Um mittels der Gleichung 3.7 die Konvergenzzeit von BGP abschätzen zu können, müssen noch die Werte für die Fehlererkennung und die Sendezeit zwischen zwei autonomen Systemen festgelegt werden. Bei der Fehlererkennung wird der schnellere Mechanismus der BFD (s. Abschnitt 3.3.1) mit einer geschätzten Dauer von 300ms berücksichtigt und die mittlere Sendezeit wird mit $\mathrm{T_V} = 20\mathrm{ms}$ angenommen.

3.4.5 Minimale BGP-Konvergenzzeit

Das Ziel dieser Abschätzung ist es, die untere Grenze der BGP-Konvergenzzeit zu finden, das bedeutet, daß das Hauptinteresse beim möglichst schnellen Bekanntgeben eines neuen Pfads liegt. Deswegen sollte dieser auch möglichst kurz sein, was allerdings ein Problem bei den Szenarien $\mathrm{t_{down}}$ und $\mathrm{t_{long}}$ darstellt. Setzt man voraus, daß man das Pfadtesten von BGP vermeiden kann, z.B. mittels EPIC, dann ist auch die Pfadlänge des kürzesten bzw. besten Pfads für die Abschätzung der BGP-Konvergenzzeit relevant. Aus diesem Grund wird im folgenden die BGP-Konvergenzzeit anhand der Länge der besten Pfade bestimmt.

Legt man die DLZ für autonome Systeme ohne Routen-Reflektoren und die MRAI-Zeitglieder mit ihren Standardwerten zugrunde, erhält man die BGP-Konvergenzzeiten, die in der zweiten Spalte von Tabelle 3.2 in Abhängigkeit der Pfadlänge dargestellt sind.

Pfadlänge in AS	MRAI	MRAI ohne I-BGP	ohne MRAI	mit RR	Metrik
0	35,8s	30,8s	0,80s	0,95s	1
1	71,2s	61,2s	1,15s	1,45s	91
2	106,5s	91,5s	1,50s	1,95s	475
3	141,9s	121,9s	1,85s	2,45s	10070
4	177,2s	152,2s	2,20s	2,95s	7478
5	212,6s	182,6s	2,55s	3,45s	1983
6	247,9s	212,9s	2,90s	3,95s	436
7	283,3s	243,3s	3,25s	4,45s	86
8	318,6s	273,6s	3,60s	4,95s	22
9	354,0s	304,0s	3,95s	5,45s	2
10	389,3s	334,3s	4,30s	5,95s	1
3,6	**163,4s**	**140,3s**	**2,06s**	**2,75s**	∅

Tabelle 3.2: Konvergenzzeit mit beiden MRAI, mit E-BGP-MRAI, ohne MRAI und mit einem Routen-Reflektor

Die Spalte “Metrik” in der Tabelle 3.2 zeigt die Häufigkeit der Längen der besten Pfade, die aus den öffentlich verfügbaren BGP-Routing-Tabellen des Betreibers Telstra [Tel05] entnommen wurden.

Basierend auf diesen realen Daten läßt sich feststellen, daß die Konvergenzzeit von BGP für den längsten besten Pfad bei rund 390s bzw. bei 6,5min liegt. Wichtet man die BGP-Konvergenzzeiten der verschiedenen Pfadlängen mit deren Häufigkeit (Spalte “Metrik”), erhält man die Mittelwerte, die in der letzten Zeile angegeben sind und die einen Anhaltspunkt für die mittlere Dauer der BGP-Konvergenzzeit bieten.

Folgt man der Empfehlung aus [RLH06] und deaktiviert den MRAI für I-BGP, erzielt man die BGP-Konvergenzzeiten, die in der dritte Spalte von Tabelle 3.2 zu sehen sind. Bei diesen läßt sich eine geringfügige Verbesserung feststellen, die aber noch nicht die untere Grenze repräsentieren.

Dafür muß auch der MRAI für E-BGP ausgeschaltet werden, womit sich die BGP-Konvergenzzeiten der zwei verbleibenden Spalten ergeben. Mittels dieser Maßnahme erreicht man Konvergenzzeiten von BGP, die im Bereich von einigen Sekunden liegen. Bei der fünften Spalte von Tabelle 3.2 wurde noch pro durchlaufenem autonomen System ein Routen-Reflektor berücksichtigt, was zu den leicht vergrößerten BGP-Konvergenzzeiten im Vergleich zur vorherigen Spalte führt.

Die restlichen Bestandteile, die in der Abschätzung der BGP-Konvergenzzeit berücksichtigt wurden, sind entweder durch die physikalischen Eigenschaften der Verbindungen oder der Leistungsfähigkeit der BGP-Router gegeben. Da diese nicht ohne weiteres verändert werden können, re-

präsentieren die BGP-Konvergenzzeiten aus den letzten beiden Spalten von Tabelle 3.2 die untere Grenze für die Verteilung von BGP-Routing-Informationen im heutigen Internet.

Durch das Deaktivieren der MRAI gelangt die BGP-Konvergenzzeit in einen Bereich, der den Anforderungen der Datenübertragung von Verkehr mit garantierter Dienstqualität entspricht. Allerdings ist die Frage noch nicht beantwortet, ob der MRAI reduziert bzw. im besten Fall sogar deaktiviert werden kann. Diese Fragestellung wird im folgenden Abschnitt erörtert.

3.5 Beschleunigung der BGP-Konvergenz

Die Abschätzung der DLZ in den letzten Abschnitten hat gezeigt, daß der MRAI den entscheidenden Anteil an der Größenordnung der DLZ aufweist. Vor allem wenn die Standardwerte des MRAI benutzt werden, haben die restlichen Bestandteile nur noch einen geringen Einfluß auf die DLZ. Dies ändert sich allerdings, wenn man die Werte des MRAI reduziert bzw. den MRAI komplett ausschaltet. Wozu die Reduzierung oder Deaktivierung des MRAI führt, wird im folgenden anhand von Simulationen untersucht.

Bei den bisherigen Arbeiten, wie z.B. [GP01, NCM02, PAM+05, CDK+05], wurde das Konvergenzverhalten von BGP nur eingeschränkt untersucht. So wurden meist vollvermaschte Topologien eingesetzt und nur ein oder zwei der vier Fehlerszenarien betrachtet. Deswegen werden bei den folgenden Simulationen aller vier Fehlerszenarien Internet-Topologien verwendet. Desweiteren wird auch der Einfluß der internen Topologie eines autonomen Systems bei den simulativen Untersuchungen berücksichtigt.

3.5.1 Einfluß des Zeitglieds MRAI

Der MRAI ist ein fester Bestandteil des BGP-Standards und wurde in BGP aufgenommen, um die Nachrichtenanzahl während der Reaktion von BGP auf eine Topologieänderung zu reduzieren. Dabei legt der MRAI die Zeit fest, die vergehen muß, bevor die nächste BGP-UPDATE-Nachricht, die eine Routing-Information für dasselbe Präfix enthält, abgeschickt werden kann. Laut [RLH06] wird der MRAI angewendet auf das Versenden von BGP-UPDATE-Nachrichten getrennt nach Präfix und nach Nachbar-AS.

Die Anzahl der versendeten BGP-UPDATE-Nachrichten kann mittels des MRAI nur reduziert werden, wenn ein BGP-Router innerhalb einer MRAI-Periode mehrere BGP-UPDATE-Nachrichten empfängt, die zu einem besseren Pfad sowie zu ein und demselben Präfix führen. Da ein BGP-Router nur eine BGP-UPDATE-Nachricht verschickt, wenn er einen neuen besten Pfad ermittelt hat und ebenfalls den MRAI einsetzt, kann ein BGP-Router mehrere BGP-UPDATE-Nachrichten, auf die er einen MRAI anwenden kann, nur von verschiedenen BGP-Routern erhalten.

Daraus resultiert, daß die optimale Größe des MRAI den Laufzeitunterschieden zwischen den von verschiedenen BGP-Routern empfangenen BGP-UPDATE-Nachrichten entspricht. Wird der

MRAI zu klein gewählt, werden die BGP-UPDATE-Nachrichten mit einer längeren Laufzeit erst nach einem weiteren MRAI weitergeschickt. Ein zu groß gewählter MRAI führt zu längeren Wartezeiten, die keine weitere Reduzierung der Nachrichtenanzahl bringt, aber die Konvergenzzeit erhöht.

Aufgrund dieses Verhaltens ist zu erwarten, daß ein minimaler Wert des MRAI in Abhängigkeit von der Ausbreitungsgeschwindigkeit der BGP-UPDATE-Nachrichten existiert, was mittels der folgenden Simulation untersucht werden soll.

3.5.2 Simulationsmodell

Um die Simulationsergebnisse verstehen und bewerten zu können, folgt eine kurze Beschreibung des verwendeten Simulators und der untersuchten Szenarien.

3.5.2.1 Simulator SSFnet

Als Simulationswerkzeug wurde das *Scalable Simulation Framework* (SSFnet) [ssf06] ausgewählt, da man damit das BGP-Zeitverhalten auf Paketebene untersuchen kann und es eine in [Pre03] validierte BGP-Implementierung besitzt, die auch von allen anderen paketbasierten Simulatoren für BGP benutzt wird.

Neben den grundlegenden Eigenschaften ist es mit SSFnet möglich, I-BGP, E-BGP, den MRAI, das RFD sowie Policies zu konfigurieren. Aufgrund der frei verfügbaren Quellen der BGP-Implementierung können diese auch erweitert werden, was in dieser Arbeit für die Erweiterung von SSFnet um die grundlegende Funktionalität von EPIC genutzt wurde.

Desweiteren kann in SSFnet die Verarbeitungszeit eines Knotens modelliert werden. Dabei besitzt jeder Knoten eine Warteschlange, die nach dem *First In First Out* (FIFO) Prinzip funktioniert und in die die ankommenden BGP-Nachrichten einsortiert werden. Die in dieser Warteschlange enthalten BGP-Nachrichten werden sequentiell abgearbeitet, und die Verarbeitungszeit unterliegt einer Gleichverteilung, deren Maximalwert festgelegt werden kann. Mit dieser Implementierung kann neben der Wartezeit in der Warteschlange auch die Zeit zum Verarbeiten der BGP-UPDATE-Nachrichten modelliert werden.

3.5.2.2 Verwendete Simulationsszenarien

Als Internet-Topologien wurden die in Abschnitt 2.4.3 vorgestellten Topologien, mit jeweils 20, 50, 100 und 200 autonomen Systemen verwendet. Die Übertragungszeit der Verbindungen zwischen den autonomen Systemen wurde zufällig nach einer Gleichverteilung zwischen $5\mathrm{ms}$ und $15\mathrm{ms}$ ausgewählt. Zur Vereinfachung der Simulationen besteht ein autonomes System aus einem BGP-Router, wobei die Untersuchungen in Abschnitt 3.5.3.4 eine Ausnahme bilden.

Eine weitere Vereinfachung ist die Bestimmung der besten Pfade, ohne die Policies zu berücksichtigen. Das führt dazu, daß alle Verbindungen für das BGP-Routing zur Verfügung stehen.

Somit ist das entscheidende Kriterium für die Suche nach den besten Pfaden die Länge der Pfade in AS-Hops.

Um alle Fälle der BGP-Reaktion auf Topologieänderungen zu berücksichtigen, werden alle vier typischen Fehlerszenarien, wie sie in Abschnitt 3.2.2 beschrieben wurden, untersucht. Fast alle bisher veröffentlichten Simulationen des BGP-Konvergenzverhaltens konzentrieren sich auf das Szenario $\mathrm{t_{down}}$, wo ein Präfix nicht mehr erreicht werden kann, und einige berücksichtigen das Komplementärereignis $\mathrm{t_{up}}$, aber keine simulative Untersuchung hat bisher alle Szenarien in ihre Betrachtung mit einbezogen.

Da SSFnet nur das Starten und Stoppen aller BGP-Verbindungen eines autonomen Systems beherrscht, wurden die Fehlerbilder durch zusätzliche autonome Systeme nachgebildet. Um die Nichterreichbarkeit eines Präfixes zu simulieren, wurde ein zusätzliches autonomes System zu der jeweiligen Topologie hinzugefügt und der Ausfall und die Reparatur dieses autonomen Systems simuliert. Dabei wurde für jede Simulation ein autonomes System ausgewählt, und an dieses wurde das zusätzliche autonome System, wie in Abbildung 3.10 (a) dargestellt ist, angebunden. Dies wurde für alle autonomen Systeme wiederholt und die Ergebnisse wurden über alle autonomen Systeme gemittelt.

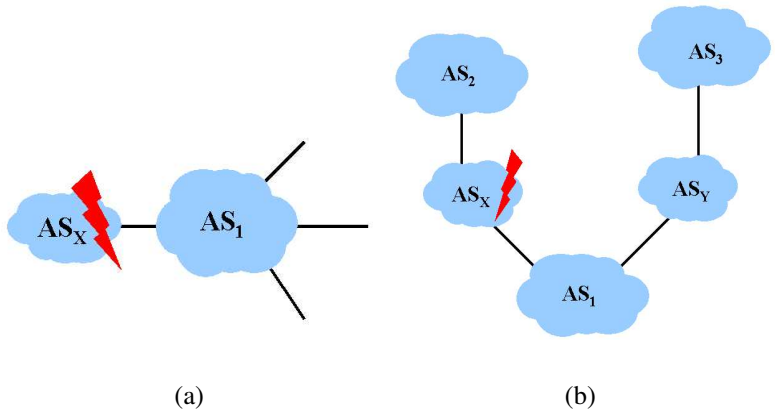


Abbildung 3.10: Modellierung der Fehlerbilder in SSFnet (a) t_{down}; (b) t_{long}

Für die Untersuchung des Umschaltens auf einen Ersatzpfad wurde pro Simulation ein autonomes System, das zweifach angebunden ist (Stub-AS), ausgewählt, und die beiden Verbindungen wurden aufgetrennt. Um die Auswahl der besten Pfade nicht zu beeinflussen, wurden die beiden Pfade des jeweiligen Stub-AS jeweils mit einem zusätzlichen autonomen System verlängert (wie $\mathrm{AS_x}$und $\mathrm{AS_y}$ in Abbildung 3.10 (b)). Im Anschluß wurden die Simulationen in zwei Läufe aufgeteilt. In einem Lauf fällt jeweils eines der zusätzlichen autonomen Systeme aus und wird wieder eingeschaltet. Im zweiten Lauf passiert dasselbe mit dem anderen zusätzlichen autonomen System. Dieses Vorgehen wurde für alle Stub-AS wiederholt und der Mittelwert der Ergebnisse aller Fehlerfälle ermittelt.

Darüber hinaus wird der Nutzdatenverkehr bei den Simulationen nicht betrachtet. Dieser kann dazuführen, daß die verfügbare Kapazität einer Verbindung nicht ausreicht und deswegen eine

BGP-Nachricht nicht übertragen wird. Da aber die BGP-Nachrichten mittels TCP transportiert werden, ist diese Übertragung gegen einen Paketverlust abgesichert. Das einzige, was passieren kann, ist eine zusätzliche Verzögerung beim Versenden der BGP-Nachrichten. Aus diesem Grund wird der Einfluß der Datenübertragung der Nutzdaten auf das BGP-Routing vernachlässigt.

3.5.2.3 Parameterauswahl

Bei allen Simulationen wurde der Wert des MRAI mit 0; 1; 3; 5; 8; 10; 20; 30 Sekunden variiert. Als weiterer Parameter wurde die maximale Verarbeitungszeit verändert. Diese wurde bei der Untersuchung der Abhängigkeit des BGP-Konvergenzverhaltens von der Verarbeitungszeit auf die Werte 0; 0,1; 0,2; 0,5; 1 Sekunden eingestellt. Bei allen anderen Simulationen war die maximale Verarbeitungszeit mit 0,2s fest vorgegeben.

Aufgrund der einfachen Kriterien für die Suche nach den besten Pfaden von BGP wurde als weiteres Entscheidungskriterium für mehrere gleich gute Pfade die Ankunftszeit der entsprechenden Routing-Information ausgewählt, wobei die Routing-Information, die als erstes empfangen wurde, bevorzugt wird.

3.5.3 Optimale Einstellung des MRAI

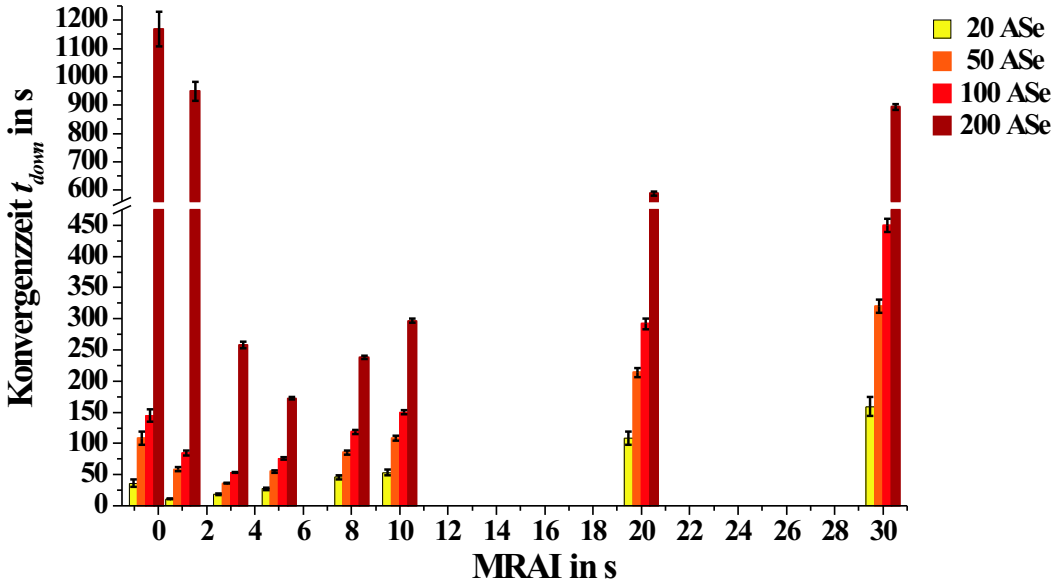
3.5.3.1 Einfluß der Netzgröße

Bei der ersten Simulationsreihe wird der Einfluß des Zeitglieds MRAI und der Größe der Internet-Topologie untersucht. Dazu wurden für alle vier Internet-Topologien und für alle MRAI-Werte die Konvergenzzeit und die Anzahl der versendeten Nachrichten ermittelt. Die Ergebnisse für den Fall, daß ein Präfix nicht mehr erreichbar ist (Szenario $\mathrm{t_{down}}$), sind in Abbildung 3.11 dargestellt.

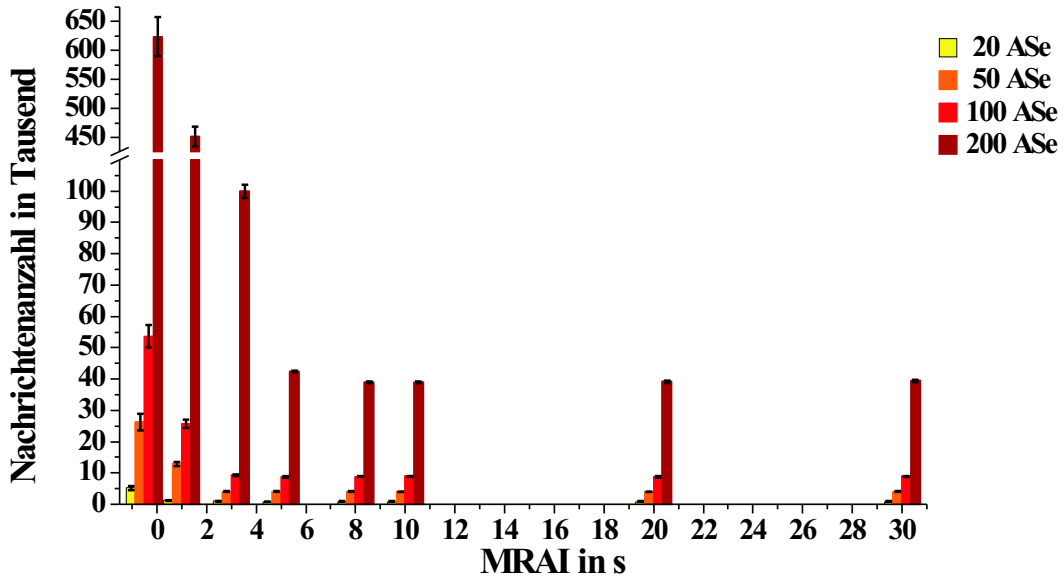
Prinzipiell verhalten sich die Ergebnisse für alle vier Internet-Topologien sehr ähnlich. Bei ausgeschaltetem MRAI zeigen sich die Auswirkungen des BGP-Pfadtestens am deutlichsten, da hierbei eine große Anzahl an Pfaden durchprobiert wird und dabei sehr viele Nachrichten generiert werden. Die stärker ansteigende Nachrichtenanzahl mit wachsender Größe der vier Internet-Topologien ist mit der größeren Zunahme der möglichen Pfade zu einem Ziel im Vergleich zur Steigerung der AS-Anzahl zu erklären. Aufgrund der extrem großen Nachrichtenanzahl sind auch sehr lange Konvergenzzeiten festzustellen.

Aktiviert man den MRAI, dann reduziert sich die Anzahl der versendeten Nachrichten maximal um den Faktor sechzehn, und damit wird das Verbesserungspotential des MRAI sehr deutlich. Allerdings bewirkt die Erhöhung des MRAI nur bis zu einem gewissen Punkt eine Reduzierung der Nachrichtenanzahl, wie es in Abschnitt 3.5.1 erwartet wurde.

Die Reduzierung der BGP-UPDATE-Nachrichten führt auch zu einer Verringerung der BGP-Konvergenzzeit. Ab dem Punkt, wo der MRAI die Nachrichtenanzahl nicht weiter reduzieren kann, beginnt die Konvergenzzeit linear zu steigen, da der MRAI hierbei nur unnötig die Weiterleitung der BGP-UPDATE-Nachrichten verzögert.



(a)



(b)

Abbildung 3.11: Konvergenzzeit (a) und Nachrichtenanzahl (b), wenn ein Präfix nicht mehr erreichbar ist

Zusammenfassend kann man feststellen, daß die Simulationsergebnisse für das Szenario t_{down} dem erwarteten Verhalten aus Abschnitt 3.5.1 entsprechen. Genau wie die Ergebnisse in [GP01, NCM02] zeigen sie, das es im Szenario t_{down} einen optimalen Wert gibt, wo die BGP-Konvergenzzeit minimal ist.

Ist das ausgefallene Präfix wieder erreichbar, wird das BGP-Konvergenzverhalten nicht vom Pfadtesten beeinflußt. Dessen Simulationsergebnisse sind in Abbildung 3.12 zu sehen.

Hierbei steigt die Konvergenzzeit linear mit größer werdendem MRAI an, und die Nachrichtenanzahl bleibt nahezu konstant. Im Gegensatz zum Szenario t_{down} verzögert der MRAI nur die Konvergenz von BGP und führt auch zu keiner Verringerung der Nachrichtenanzahl. Desweiteren bleibt festzustellen, daß alle Werte drastisch kleiner ausfallen. So bewegt sich die Konvergenzzeit im Bereich von einer MRAI-Periode, und die Nachrichtenanzahl liegt im Mittel etwas über dem Zweifachen der Verbindungsanzahl. Das zeigt, daß ein autonomes System im Mittel pro Interdomain-Verbindung zwei BGP-UPDATE-Nachrichten erhält.

3.5.3.2 Konvergenzverhalten beim Umschalten auf einen Ersatzpfad

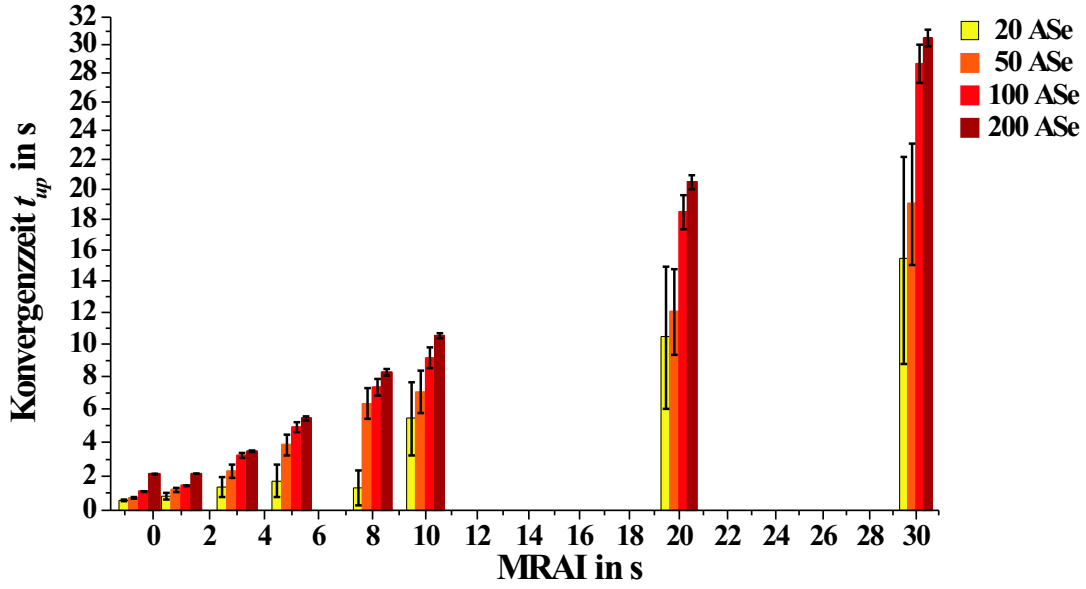
Bevor auf die Simulationsergebnisse, die das Umschalten auf einen Ersatzpfad untersuchen, eingegangen werden kann, muß noch eine Besonderheit erklärt werden, die dem Simulationsmodell mit zwei zusätzlichen autonomen Systemen geschuldet ist (s. Abschnitt 3.5.2).

In den Abbildungen, die die Ergebnisse der Szenarien t_{long} und t_{short} darstellen, sind immer zwei Werte angegeben. Die Werte, die als "vollständig" bezeichnet werden, zeigen alle versendeten BGP-UPDATE-Nachrichten und die damit verbundene BGP-Konvergenzzeit. Allerdings können diese Werte durch die Nichterreichbarkeit bzw. Reaktivierung des zusätzlichen autonomen Systems aus dem Simulationsmodell verfälscht werden.

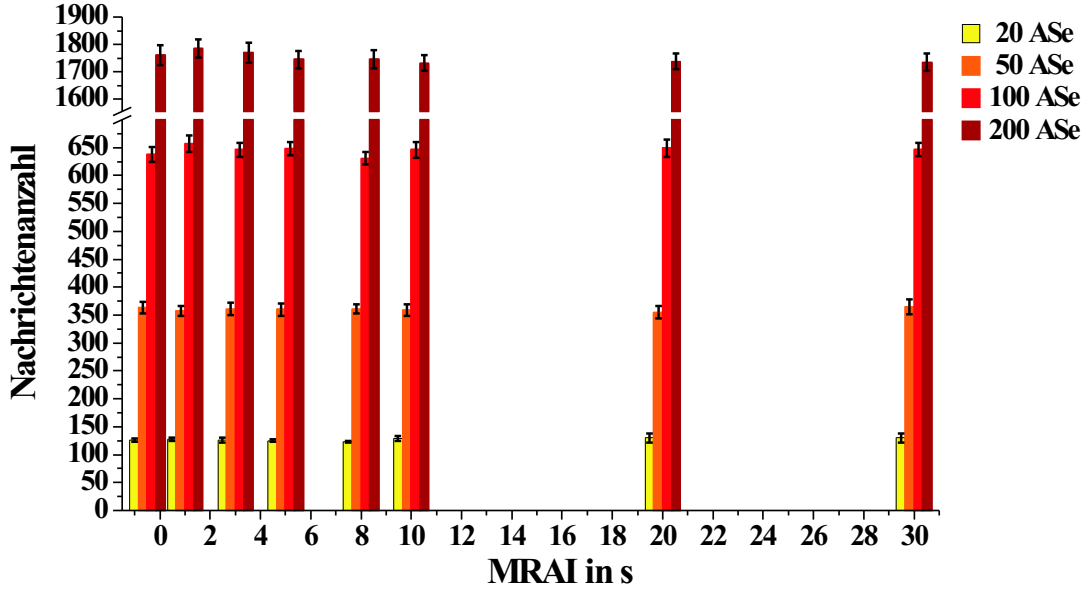
Deswegen ist in allen betroffenen Abbildungen noch ein zweiter Wert, der die Bezeichnung "korrigiert" trägt, dargestellt. Dabei werden alle BGP-UPDATE-Nachrichten, die von den zwei zusätzlichen autonomen Systemen empfangen bzw. gesendet werden, herausgefiltert. Somit veranschaulichen die "korrigierten" Werte nur die BGP-UPDATE-Nachrichten, die den Ersatzpfad (t_{long}) oder den reparierten Pfad (t_{short}) bekanntgeben.

Im Fall des Ausfalls einer Interdomain-Verbindung stellen die "korrigierten" Werte das Konvergenzverhalten von BGP beim Umschalten auf einen Ersatzpfad nach dem Ausfall des besten Pfads dar. In den "vollständigen" Ergebnissen ist zusätzlich die Routing-Information, daß das Präfix des ausgefallenen zusätzlichen autonomen Systems nicht mehr erreichbar ist, enthalten.

Beim Reaktivieren des ausgefallenen autonomen Systems repräsentieren die "vollständigen" Ergebnisse nicht nur den Einfluß des Simulationsmodells. Neben der Bekanntgabe des neuen besten Pfads werden auch die Erreichbarkeit des Präfixes des zusätzlichen autonomen Systems und alle Erreichbarkeitsinformationen der beiden Nachbar-AS an das reaktivierte autonome System übertragen. Da nach der Reparatur einer Interdomain-Verbindung die über diese miteinander verbundenen autonomen Systeme gegenseitig die erreichbaren Präfixe austauschen, muß zu der "korrigierten" Nachrichtenanzahl noch zweimal die Anzahl der erreichbaren Ziele (entspricht



(a)



(b)

Abbildung 3.12: Konvergenzzeit (a) und Nachrichtenanzahl (b), wenn ein Präfix wieder erreichbar ist

der Anzahl der autonomen Systeme minus eins) addiert werden. Die restlichen Nachrichten werden zur Bekanntgabe des Präfixes des zusätzlichen autonomen Systems und der Interaktion der zusätzlichen autonomen Systeme mit deren Nachbar-AS verwendet. Allerdings ist es bei der Konvergenzzeit nicht möglich, den Einfluß des Simulationsmodells herauszurechnen, da sich die einzelnen BGP-UPDATE-Nachrichten bei der Wartezeit in den Warteschlangen und während der Verarbeitung in einem BGP-Router gegenseitig beeinflussen.

In Abbildung 3.13 sind die Simulationsergebnisse für das Szenario $\mathrm{t_{long}}$ veranschaulicht, wobei hier nur die Internet-Topologie mit 100 autonomen Systemen berücksichtigt wird.

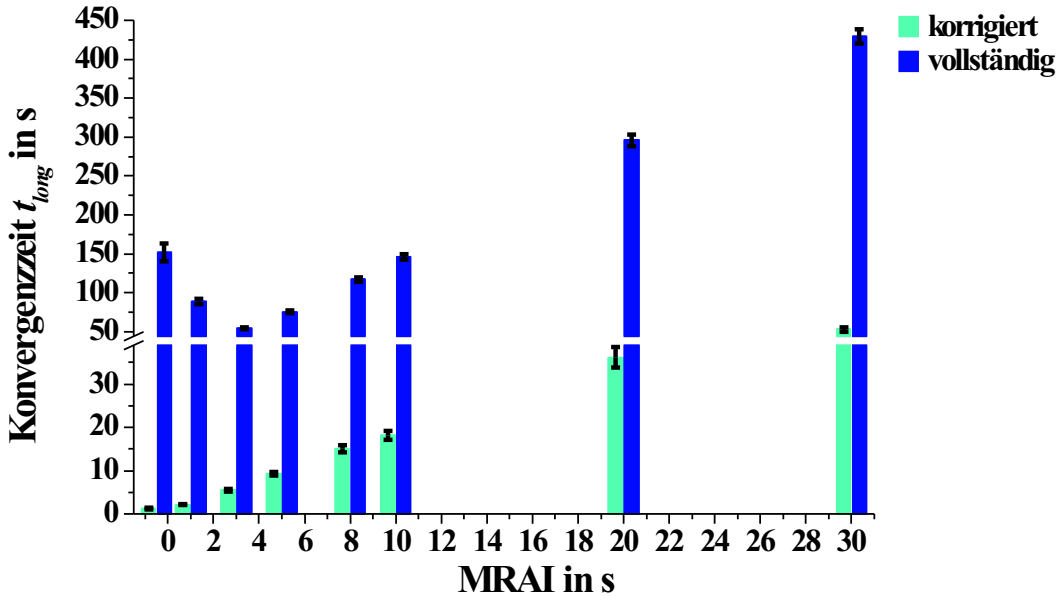
Bei der Bekanntgabe des Ersatzpfads ("korrigierte" Werte) vergrößert sich die Konvergenzzeit linear mit steigendem MRAI und die Nachrichtenanzahl erhöht sich leicht. Betrachtet man die "vollständigen" Ergebnisse, entsprechen diese fast den Werten für den Fall $\mathrm{t_{down}}$. Im Gegensatz zu den Meßergebnissen, die in Abschnitt 3.2.1 beschrieben wurden, sind die "korrigierten" Werte deutlich kleiner und zeigen auch eine andere Abhängigkeit vom MRAI. Trotzdem stellen die "korrigierten" Werte die BGP-Reaktion korrekt dar. Allerdings ist der Unterschied zwischen den ungültigen Pfaden und dem Ersatzpfad aufgrund der Nichtbetrachtung der Policies relativ gering. Deswegen beendet der Ersatzpfad das Pfadtesten von BGP relativ schnell, was zu den deutlich kleineren "korrigierten" Werten bei der Nachrichtenanzahl führt, welche sich auch in wesentlich kürzeren Konvergenzzeiten niederschlagen.

Anhand der "vollständigen" Ergebnisse kann man das Konvergenzverhalten von BGP ablesen, wenn der Unterschied zwischen dem normalen Pfad und dem Ersatzpfad aufgrund der Policies deutlich größer ist. Das Konvergenzverhalten von BGP verändert sich mit zunehmendem Unterschied zwischen dem Ersatzpfad und dem besten Pfad von dem Verhalten der "korrigierten" hin zu den "vollständigen" Ergebnissen. Der schlechteste denkbare Fall, der exakt durch die "vollständigen" Werte repräsentiert wird, tritt ein, wenn der Ersatzpfad dem schlechtest möglichen Pfad entspricht. Je nachdem, ob das Pfadtesten von BGP stark ausgeprägt ist oder nicht, kann der MRAI zu einer Verbesserung der Fehlerreaktion von BGP beitragen oder diese verzögern.

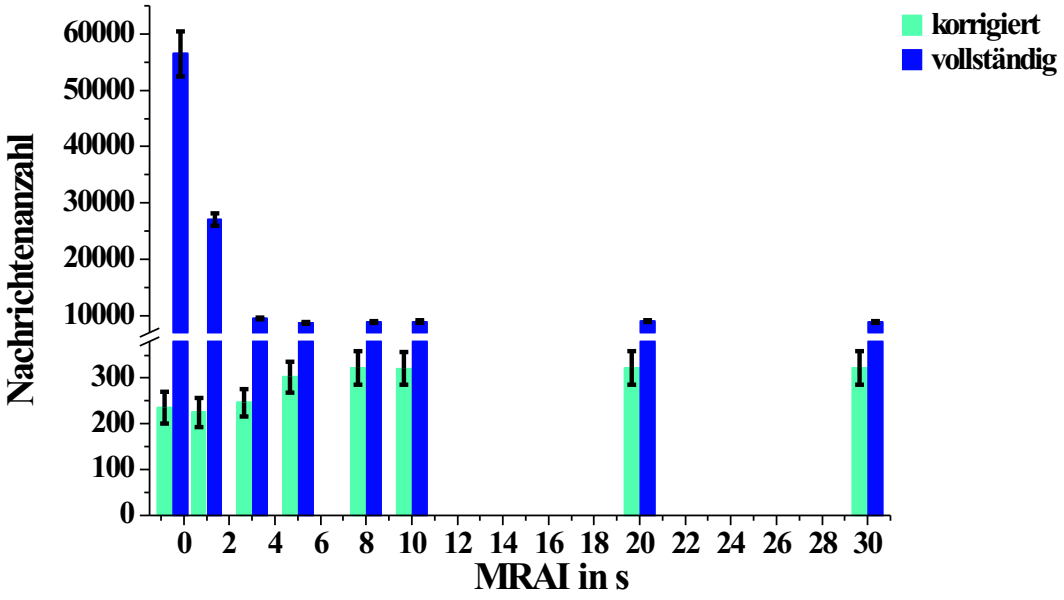
Im Gegensatz dazu hat der MRAI nur einen verzögernden Einfluß, wenn die ausgefallene Interdomain-Verbindung repariert wird und BGP auf den wieder funktionstüchtigen Pfad zurückwechselt. Die Simulationsergebnisse für diesen Fall sind in Abbildung 3.14 mit den Resultaten von $\mathrm{t_{up}}$ verglichen.

Betrachtet man die Nachrichtenanzahl, so stellt man fest, daß diese für die "korrigierte" Bekanntgabe eines neuen besten Pfads deutlich am kleinsten ist. Hier wird deutlich, daß sich der beste Pfad nur in einigen autonomen Systemen ändert und für die anderen autonomen Systeme diese Änderung irrelevant ist. Um die tatsächliche Nachrichtenanzahl zu erhalten, muß, wie schon erwähnt, die zweifache Anzahl der erreichbaren Ziele hinzuaddiert werden. Da die betrachtete Internet-Topologie aus 100 autonomen Systemen besteht, kann man ungefähr 200 Nachrichten zu den "korrigierten" Werten hinzuzählen. Somit liegen diese Ergebnisse unter den Resultaten für die Bekanntgabe eines neuen Präfixes. Genauso wie im Fall $\mathrm{t_{long}}$ steigen diese Werte, wenn sich der Ersatzpfad und der beste Pfad stärker unterscheiden.

Der Einfluß des Simulationsmodells verschiebt auch die "vollständigen" Werte der Konvergenzzeit. Dabei führt das Austauschen der erreichbaren Ziele zwischen dem wiederaktivierten auto-

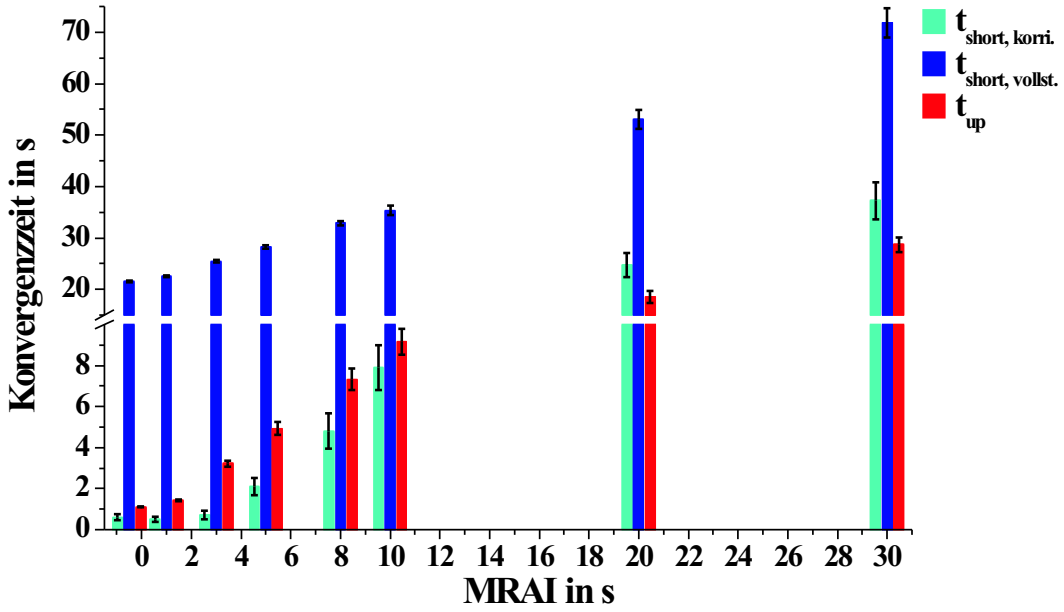


(a)

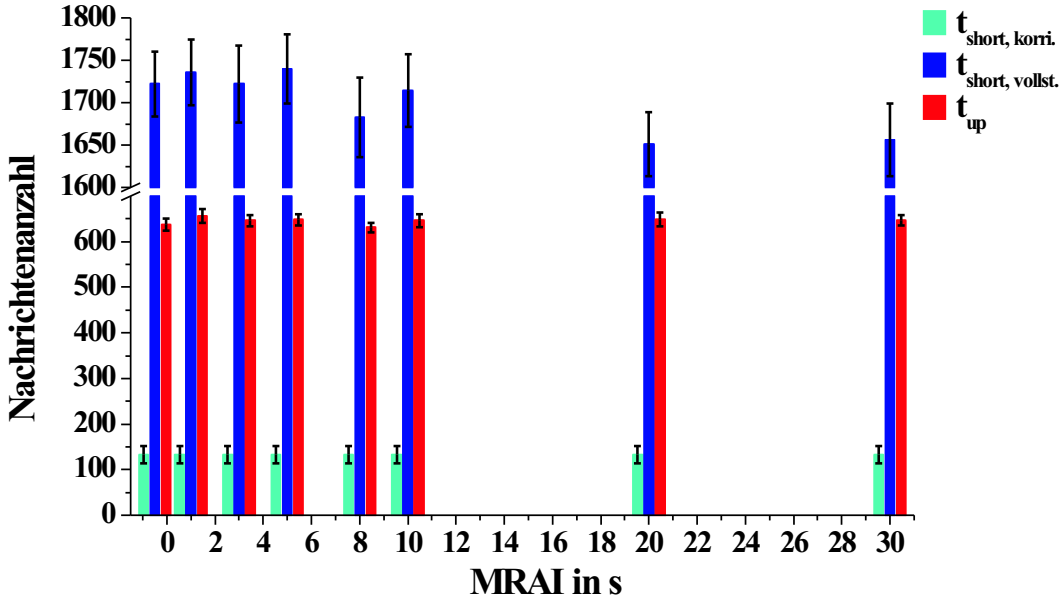


(b)

Abbildung 3.13: Konvergenzzeit (a) und Nachrichtenanzahl (b) beim Umschalten auf einen Ersatzpfad



(a)



(b)

Abbildung 3.14: Vergleich des Konvergenzverhaltens von t_{short} und t_{up} anhand der Zeit (a) und der Nachrichtenanzahl (b)

nomen System und den beiden angeschlossenen autonomen Systemen zu einem Offset von rund 20 Sekunden sowie zu einem flacheren Anstieg der Konvergenzzeit mit zunehmendem MRAI. Die Konvergenzzeiten der beiden anderen Szenarien nehmen ebenfalls mit dem MRAI zu, allerdings reduziert um den Offset und mit einem steileren Anstieg. Dabei vergrößern sich die Konvergenzzeiten der "korrigierten" Werte etwas stärker als die Ergebnisse von $\mathrm{t_{up}}$. Der Grund dafür liegt in den beteiligten autonomen Systemen. Da im Fall von $\mathrm{t_{short}}$ relativ wenige autonome Systeme an der Fehlerreaktion beteiligt sind, werden pro autonomen System mehr BGP-UPDATE-Nachrichten versandt und dadurch werden mehr BGP-UPDATE-Nachrichten durch den MRAI verzögert.

3.5.3.3 Untersuchung verschiedener Verarbeitungszeiten

Für die Untersuchung des Einflusses der Verarbeitungszeit in einem BGP-Router wurde die maximale Verarbeitungszeit variiert, wie es in Abschnitt 3.5.2.1 beschrieben wurde. Als Fehlerszenario wurde die Nichterreichbarkeit eines Präfixes in einer Internet-Topologie mit 100 autonomen Systemen ausgewählt und die dabei erzielten Resultate sind in Abbildung 3.15 dargestellt.

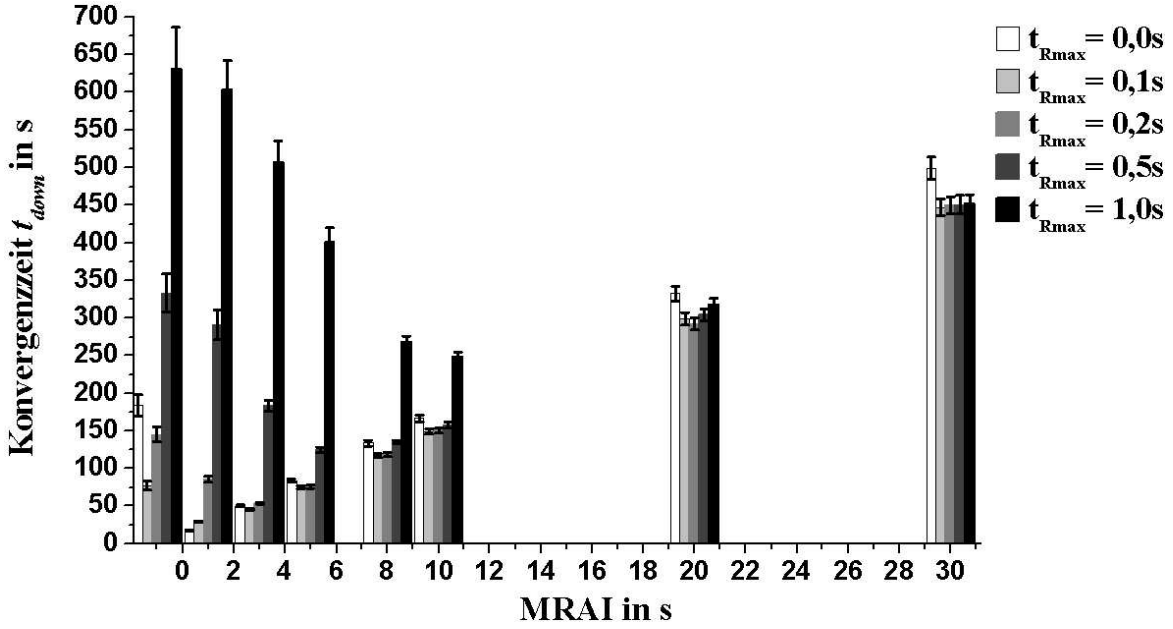
Auch diese Ergebnisse sind stark durch das Pfadtesten von BGP geprägt. Es existiert ein optimaler Wert für den MRAI, wo die BGP-Konvergenzzeit minimal ist. Allerdings ist dieses Minimum stark von der maximalen Verarbeitungszeit abhängig.

Werden die Verarbeitungszeit und der MRAI nicht berücksichtigt, schnellt die Nachrichtenanzahl sehr stark nach oben. Aktiviert man den MRAI, fällt die Nachrichtenanzahl sofort auf ihr Minimum und ändert sich nur noch unwesentlich. Die minimale Konvergenzzeit wird schon bei einem Wert des MRAI von einer Sekunde erreicht und steigt dann linear mit zunehmendem MRAI.

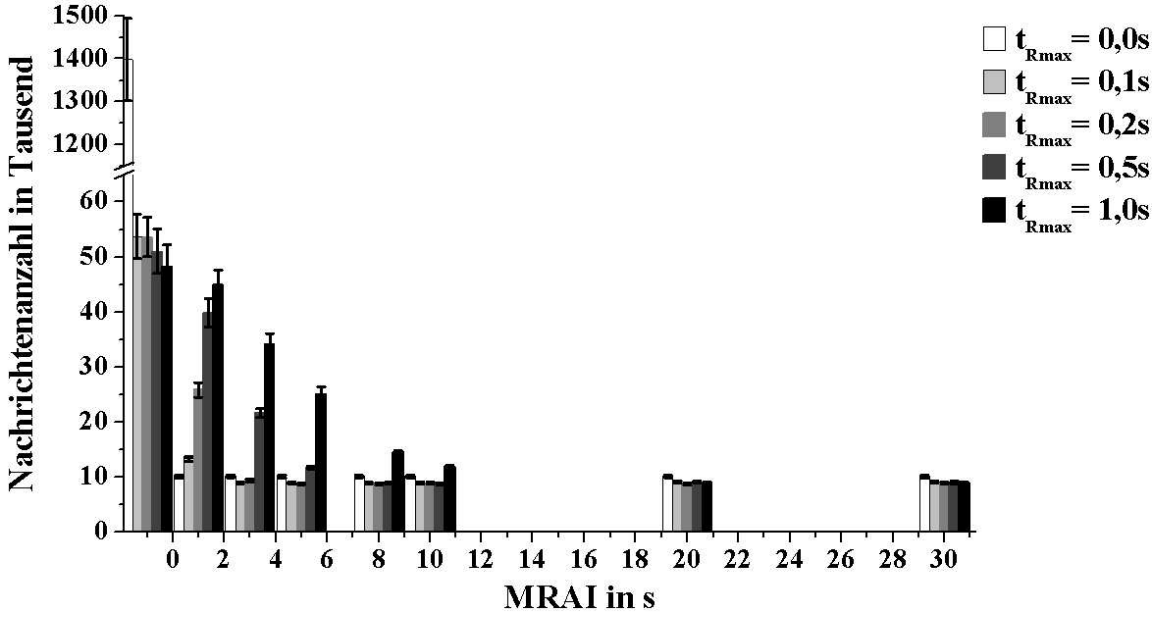
Diese Schwelle, wo die BGP-Konvergenzzeit minimal ist, vergrößert sich mit zunehmender Verarbeitungszeit. Beträgt die maximale Verarbeitungszeit eine Sekunde, erreicht die Konvergenzzeit erst bei einem MRAI von zehn Sekunden ihr Minimum, aber der typische Verlauf aufgrund des Pfadtestens von BGP bleibt erhalten. Es wird deutlich, wie stark die Verarbeitungszeit die optimale Wahl des MRAI beeinflußt.

Wie in Abschnitt 3.5.1 erläutert, erreicht man mittels des MRAI die beste Verbesserung der Konvergenzzeit von BGP, wenn der MRAI genau die Laufzeitunterschiede zwischen den relevanten Pfaden ausgleicht. Genau diese Laufzeitunterschiede werden wesentlich von der Verarbeitungszeit geprägt, da im Mittel die Verarbeitungszeiten um mindestens eine Größenordnung über den Sendezeiten liegen.

Bei der Bekanntgabe eines unbekannten Präfixes spielt die Verarbeitungszeit nur eine untergeordnete Rolle, da dabei drastisch weniger BGP-UPDATE-Nachrichten verschickt werden. Deshalb macht sich die Verarbeitungzeit nur bei MRAI-Werten von Null und einer Sekunde bemerkbar, wobei eine längere Verarbeitungszeit auch zur Erhöhung der Konvergenzzeit führt. Bei deaktiviertem MRAI beträgt die Konvergenzzeit rund eine Sekunde, wenn die Verarbeitungszeit nicht berücksichtigt wird. Aber die BGP-Konvergenzzeit erhöht sich auf vier Sekunden, wenn die maximale Verarbeitungszeit eine Sekunde beträgt und der MRAI ausgeschaltet bleibt.



(a)



(b)

Abbildung 3.15: Einfluß verschiedener Verarbeitungszeiten anhand der Zeit (a) und der Nachrichtenanzahl (b)

Für die verbleibenden Szenarien ergibt sich dieselbe Abhängigkeit, wie sie im Szenario $\mathrm{t_{down}}$ deutlich geworden ist. Das bedeutet, daß die Konvergenzzeit von BGP mit steigender Verarbeitungszeit auch ansteigt.

3.5.3.4 Einfluß der internen Struktur eines autonomen Systems auf das Konvergenzverhalten

Die meisten Untersuchungen der BGP-Konvergenzzeit benutzen die Vereinfachung, daß ein autonomes System gleich einem BGP-Router entspricht. Wie in Abschnitt 3.3.5 beschrieben, kann dies sogar dazu führen, daß ein Mechanismus nicht korrekt funktioniert, da die interne Topologie eines autonomen Systems nicht mit in die Überlegungen einbezogen wurde. Deswegen wird im folgenden anhand der Internet-Topologie mit 100 autonomen Systemen der Einfluß der internen Topologie auf das Konvergenzverhalten von BGP untersucht.

Als interne Topologien wurden in jedem autonomen System, die drei in Abschnitt 2.1.2.5 beschriebenen benutzt, wobei jeweils die Topologien mit bzw. ohne Routen-Reflektoren betrachtet wurden. Allerdings konnte der Fall, bei dem die internen Topologien aus 48 Routern bestehen und kein Routen-Reflektor verwendet wurde, nicht berechnet werden, da diese Simulation mehr als 2 GB Arbeitsspeicher benötigt, was die benutzte Hardware nicht unterstützt.

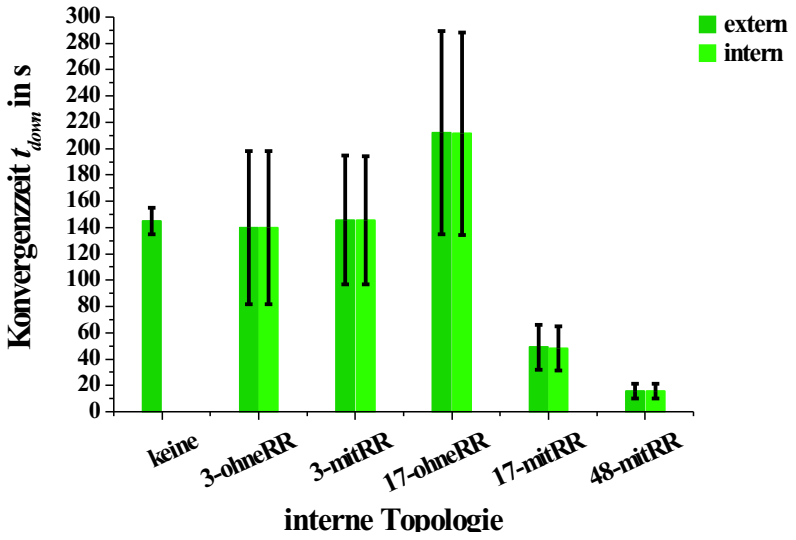
Die erzielten Ergebnisse sind in Abbildung 3.16 für das Szenario $\mathrm{t_{down}}$ mit deaktiviertem MRAI und einer maximalen Verarbeitungszeit von 200ms pro BGP-Router dargestellt.

Besteht die interne Topologie aus drei Routern (bezeichnet mit “3-ohneRR” bzw. “3-mitRR”), sind die mittleren Konvergenzzeiten vergleichbar mit den Simulationen, wo keine interne Topologie berücksichtigt werden (beschrieben mit “keine”). Wird eine größere interne Topologie mit Routen-Reflektoren betrachtet (markiert als “17-mitRR” bzw. “48-mitRR”), dann fallen die Konvergenzzeiten deutlich geringer aus. In diesen Fällen führt die viermalige Bestimmung der besten Pfade in jedem autonomen System (im Ein- und Ausgangs-Router sowie in zwei Routen-Reflektoren) zu einer Reduzierung der zwischen den autonomen Systemen (extern) versendeten BGP-Nachrichten und somit zu einer Abschwächung des BGP-Pfadtestens. Daraus resultiert in diesen beiden Fällen eine verringerte BGP-Konvergenzzeit.

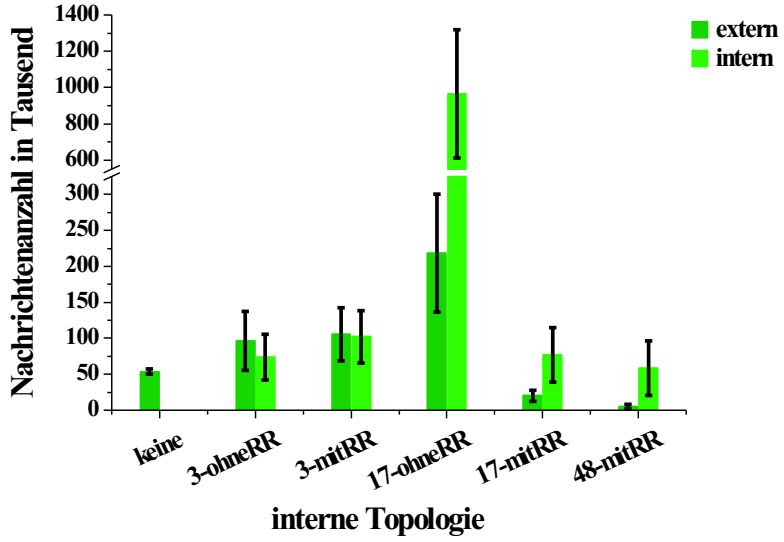
Das Gegenteil passiert bei der Verwendung der internen Topologie mit siebzehn Knoten und ohne Routen-Reflektor (bezeichnet mit “17-ohneRR”). Hierbei führt die interne BGP-Kommunikation zu einer weiteren Verzögerung beim Verteilen der BGP-Nachrichten und zu größeren Unterschieden bei den Ausbreitungszeiten. Da kein MRAI eingesetzt wird, führt diese stärkere Verteilung der Berechnung der besten Pfade zu einer Verstärkung des BGP-Pfadtestens. Das Resultat aus diesem Verhalten ist eine deutlich längere BGP-Konvergenzzeit.

Bei der Bekanntgabe eines unbekannten Präfixes (Szenario $\mathrm{t_{up}}$) zeigt sich ein etwas anderer Einfluß der internen Topologien, dessen Simulationsergebnisse in Abbildung 3.17 zu sehen sind.

Wie bei den vorherigen Simulationen kann generell festgestellt werden, daß das Nichtauftreten des BGP-Pfadtestens zu einer Reduzierung der Konvergenzzeiten und Nachrichtenanzahlen um ein bis zwei Zehnerpotenzen führt. Desweiteren kann man in diesem Szenario den Einfluß der



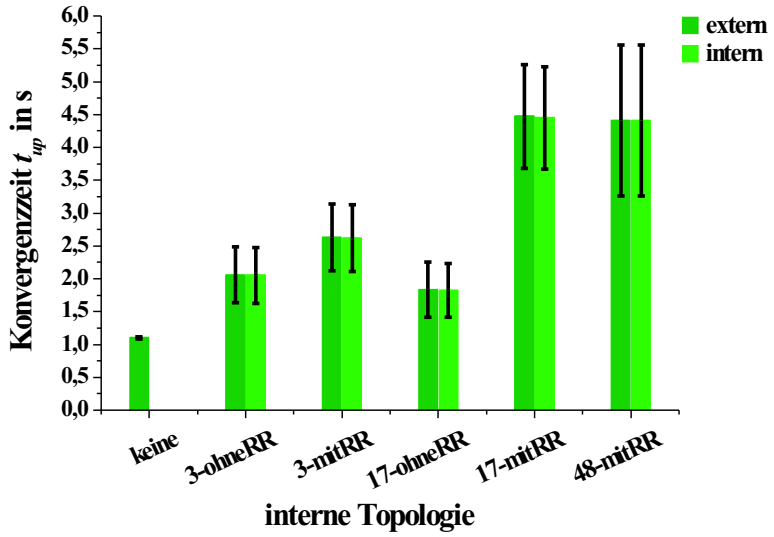
(a)



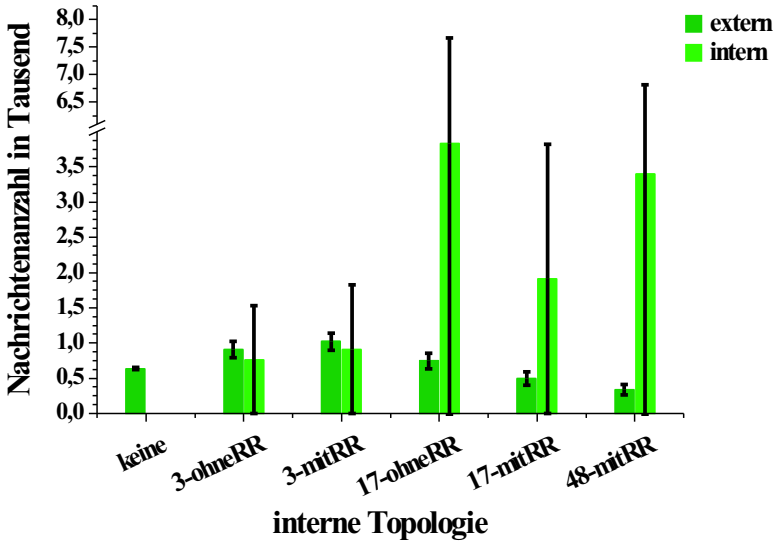
(b)

Abbildung 3.16: Einfluß der internen Topologie eines autonomen Systems anhand der Konvergenzzeit t_{down} (a) und der Nachrichtenanzahl (b)

Häufigkeit der Suche der besten Pfade beobachten. Wird keine interne Topologie berücksichtigt, wird nur einmal in jedem autonomen System der beste Pfad bestimmt. Deswegen weist dieser Fall die kleinste Konvergenzzeit auf. Betrachtet man eine interne Topologie ohne Routen-Reflektoren, wird der beste Pfad von BGP zweimal je autonomes System ermittelt. Somit ver-



(a)



(b)

Abbildung 3.17: Einfluß der internen Topologie eines autonomen Systems anhand der Konvergenzzeit t_{up} (a) und der Nachrichtenanzahl (b)

doppelt sich ungefähr die Konvergenzzeit. Kommt nur ein Routen-Reflektor pro autonomes System (Fall “3-mitRR”) zum Einsatz, ist eine Verdreifachung der Konvergenzzeit zu beobachten. Eine annähernde Vervierfachung der Konvergenzzeit läßt sich feststellen, wenn mehrere Routen-Reflektoren in jedem autonomen System benutzt werden, und somit viermal die Suche der besten

Pfade von BGP ausgeführt wird.

Auch bei der Anzahl der versendeten BGP-Nachrichten sind die Einflüsse der internen Topologien feststellbar. Da die Verteilung der BGP-Nachrichten per I-BGP durch die Größe der internen Topologie und den Einsatz der Routen-Reflektoren beeinflußt wird, kann man feststellen, daß die Anzahl der intern verschickten BGP-Nachrichten zunimmt, je mehr I-BGP Verbindungen benötigt werden. Gleichzeitig führt die häufigere Suche der besten Pfade in jedem autonomen System zu einer Reduzierung der extern übermittelten BGP-Nachrichten (via E-BGP).

Diese erzielten Werte der BGP-Konvergenzzeit bestätigen den Einfluß der internen Topologie auf die Abschätzung der BGP-Konvergenzzeit aus Abschnitt 3.4.4. Die Verzögerung, die innerhalb eines autonomen Systems aufgrund der I-BGP-Kommunikation entsteht, führt zu einer deutlichen Erhöhung der Konvergenzzeit von BGP. Auch läßt sich eine vergrößerte BGP-Konvergenzzeit feststellen, wenn Routen-Reflektoren eingesetzt werden.

Da diese Ergebnisse die Ausbreitung der BGP-UPDATE-Nachrichten auf dem besten Pfad bei ausgeschaltetem MRAI repräsentieren, bestätigen die Werte der BGP-Konvergenzzeit die Abschätzung der unteren Grenze in Abschnitt 3.4.5.

Die Aussagen der Szenarien $\mathrm{t_{down}}$ und $\mathrm{t_{up}}$ werden auch durch die Simulationsergebnisse der beiden verbleibenden Szenarien bestätigt, wobei sonst keine nennenswerten Unterschiede auftreten.

3.5.4 Bewertung der erzielten Ergebnisse

Anhand der diskutierten Simulationsergebnisse läßt sich feststellen, daß mittels des MRAI nur bei Szenarien, in denen Pfade aufgrund eines Fehlers nicht mehr benutzbar sind und damit das BGP-Pfadtesten auftritt, die Anzahl der versendeten BGP-UPDATE-Nachrichten nachweisbar reduziert werden kann. Wie erwartet, existiert ein optimaler Wert für den MRAI, wo die BGP-Konvergenzzeit minimiert wird. Allerdings ist dieser Wert von der Verarbeitungsdauer der BGP-UPDATE-Nachrichten in den BGP-Routern abhängig. Da diese aufgrund der Auslastung der BGP-Router stark schwanken kann und die Performanz der BGP-Router ebenfalls große Unterschiede aufweist, ist das Aufstellen einer Regel für das richtige Konfigurieren des MRAI nicht realisierbar. Außerdem gilt es dabei zu berücksichtigen, daß ein zu klein eingestellter MRAI die Nachrichtenanzahl drastisch erhöht. Aus diesem Grund ist der Einsatz des MRAI auch unumgänglich, um die Auswirkungen des BGP-Pfadtestens innerhalb der Leistungsfähigkeit der BGP-Router zu halten.

Im Gegensatz zu vielen anderen Simulationen wurde auch die BGP-Konvergenz bei der Bekanntgabe eines bisher unbekannten Präfixes bzw. eines besseren Pfads untersucht. In diesen Fällen verzögert der MRAI die Konvergenzzeit von BGP, ohne sonstige Vorteile aufzuweisen.

Ein wichtiger Aspekt, der hier erstmalig betrachtet wurde, ist der Einfluß der internen Topologie eines autonomen Systems auf die BGP-Konvergenzzeit. Dabei konnte gezeigt werden, daß eine interne Topologie aufgrund der Verzögerung beim Austausch der BGP-Routing-Informationen innerhalb eines autonomen Systems die Nachrichtenanzahl deutlich erhöht. Tritt in einem Szenario das BGP-Pfadtesten nicht auf, wird auch die zusätzliche Verarbeitungszeit der BGP-Router

innerhalb eines autonomen Systems und der gegebenfalls eingesetzten Routen-Reflektoren sichtbar. Die dabei erzielten Ergebnisse bestätigen die Annahmen, die bei der Abschätzung der BGP-Konvergenzzeit getroffen wurden, und die simulativ ermittelten Konvergenzzeiten von BGP entsprechen der Vorhersage der unteren Grenze für eine mittlere Pfadlänge.

Der Hauptvorteil bei der Verwendung von Internet-Topologien in den Simulationen besteht darin, daß die Simulationsergebnisse auf das Verhalten von BGP im realen Internet übertragen werden können (s. Abschnitt 2.4.3). Allerdings müssen dabei die Einflüsse des Simulationsmodells berücksichtigt werden.

Für die Szenarien $\mathrm{t_{up}}$und $\mathrm{t_{short}}$ kann davon ausgegangen werden, daß das BGP-Routing im realen Internet sich sehr ähnlich verhält. Was sich mit der wesentlich größeren Anzahl an autonomen Systemen definitiv ändert, ist die Gesamtanzahl der versendeten BGP-UPDATE-Nachrichten. Allerdings dürfte sich die mittlere Anzahl der verschickten BGP-UPDATE-Nachrichten pro BGP-Router nur marginal ändern, da die verwendeten Internet-Topologien die Länge der besten Pfade im realen Internet exakt nachbilden. Damit ist zu erwarten, daß die Verteilung der Routing-Informationen über einen neuen besseren Pfad oder ein bisher unbekanntes Präfix sich genauso wie in den Simulationen verhalten.

So einfach stellt sich die Situation bei den Szenarien, bei denen das BGP-Pfadtesten auftritt, nicht dar. Hier gilt es zu beachten, daß in den Simulationen die Anzahl der möglichen Pfade wesentlich größer ist, da sie nicht von Policies eingeschränkt werden. Damit steigen die Nachrichtenanzahl und die Konvergenzzeit in den Simulationen mit einer höheren AS-Anzahl überproportional stark an. Nichtdestotrotz kann davon ausgegangen werden, daß aufgrund des MRAI in Abhängigkeit von der Ausbreitungsdauer der BGP-UPDATE-Nachrichten ein Minimum bei der BGP-Konvergenzzeit existiert. Genau wie in den Simulationen wird die minimale BGP-Konvergenzzeit aufgrund der optimalen Einstellung des MRAI aber nicht annähernd in die Nähe der theoretischen unteren Grenze gelangen. Um dies zu erreichen, muß das Pfadtesten von BGP vermieden werden, denn nur dann ist der Einsatz des MRAI nicht mehr zwingend notwendig.

3.5.5 Einsatz von EPIC

Wie in Abschnitt 3.3.5 beschrieben und in [CDK$^+$05] bewiesen wurde, ermöglicht EPIC, das BGP-Pfadtesten zu verhindern, indem in jeder BGP-UPDATE-Nachricht die Ursache der Topologieänderung mitgeführt wird. Ist ein Präfix nicht mehr erreichbar (Szenario $\mathrm{t_{down}}$), werden ausschließlich BGP-UPDATE-Nachrichten versandt, die das betreffende Präfix nicht erreichbar deklarieren. Damit erfährt jeder BGP-Router nach Erhalt der ersten BGP-UPDATE-Nachricht den Grund für die Topologieänderung und kann alle ungültigen Pfade sofort löschen. Dieses Versenden der BGP-UPDATE-Nachrichten entspricht exakt dem Fluten des Netzes. Somit ist die Anzahl der verschickten BGP-UPDATE-Nachrichten gleich der Anzahl der Interdomain-Verbindungen plus der Anzahl der doppelt versendeten BGP-UPDATE-Nachrichten.

Für die Konvergenzzeit ist die Laufzeit der BGP-UPDATE-Nachrichten ein entscheidender Faktor, wobei hier die Sendezeit über den kürzesten Pfad zum am weitesten entfernten autonomen

System betrachtet werden muß.

Fällt eine Interdomain-Verbindung aus und BGP findet einen gültigen Ersatzpfad, dann verhindert EPIC ebenfalls das BGP-Pfadtesten. Dieser Fehlerfall reduziert sich auf die Bekanntgabe des Ersatzpfads als neuen besten Pfad, da wie bei der Nichterreichbarkeit eines Präfixes nach dem Empfang der ersten BGP-UPDATE-Nachricht alle ungültigen Pfade gelöscht werden. Dabei gelten die Abschätzungen der Konvergenzzeit, wie sie im Abschnitt 3.4.5 aufgestellt wurden.

Betrifft der Fehler mehr als ein Präfix, gelten dieselben Aussagen, da für jede Routing-Änderung jeweils eigene BGP-UPDATE-Nachrichten versendet werden. Für den Fall, daß mehrere Fehler, die dasselbe Präfix betreffen, gleichzeitig auftreten, werden diese Informationen ebenfalls getrennt mittels verschiedener BGP-UPDATE-Nachrichten verteilt. Somit kann ein BGP-Router mittels EPIC auch mehrere Fehlerquellen erkennen und sofort alle ungültigen Pfade löschen. Es erhöht sich allerdings damit die Anzahl der gesendeten BGP-UPDATE-Nachrichten, und die Konvergenzzeit vergrößert sich in Abhängigkeit des Zeitunterschieds zwischen den einzelnen Fehlern. Da ein Router-Ausfall auch als Ausfall der Verbindungen, mit denen der Router mit seinen Nachbarn verbunden ist, verstanden werden kann, gilt dasselbe für den Ausfall eines Routers. Damit stellt EPIC eine wirkungsvolle Erweiterung von BGP dar, die das Pfadtesten von BGP vermeidet und somit die Konvergenzzeit beschleunigt.

3.5.5.1 Fehlerreaktion von EPIC

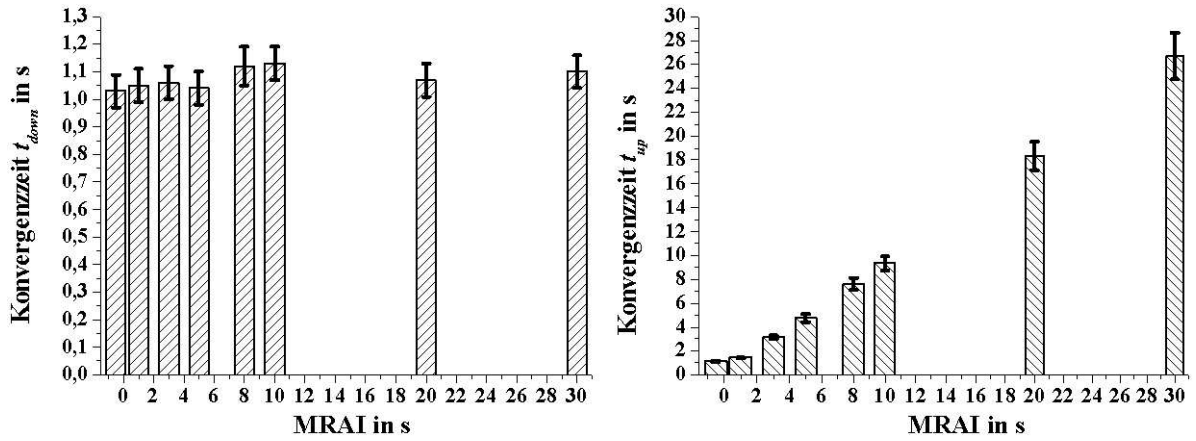
In diesem Abschnitt sollen die im letzten Abschnitt prognostizierten Eigenschaften von EPIC per Simulation validiert werden. Abbildung 3.18 zeigt die Simulationsergebnisse für die Szenarien, daß ein Präfix nicht mehr (Szenario $\mathrm{t_{down}}$) und nach einer gewissen Zeit wieder erreichbar ist (Szenario $\mathrm{t_{up}}$) anhand der Internet-Topologie mit 100 autonomen Systemen.

Bei der Nichterreichbarkeit eines Präfixes werden in allen Simulationen genau 473 BGP-UPDATE-Nachrichten verschickt, d.h. daß die Nachrichtenanzahl unabhängig vom MRAI ist. Desweiteren bestätigt dies, daß EPIC die BGP-UPDATE-Nachrichten, die nur Routing-Informationen im Withdrawn-Teil enthalten, flutet. Aufgrund dessen liegt die Konvergenzzeit für alle betrachteten MRAI-Werte auch im Bereich einer Sekunde. Der MRAI hat überhaupt keinen Einfluß auf das Konvergenzverhalten von EPIC im Szenario $\mathrm{t_{down}}$.

Ist das ausgefallene Präfix wieder erreichbar, ist das Verhalten von EPIC gleich dem von BGP. Das bedeutet, daß die Konvergenzzeit linear mit größer werdendem MRAI ansteigt und die Nachrichtenanzahl nahezu konstant ist. Damit verzögert der MRAI nur die Konvergenz.

In Abbildung 3.19 sind die Simulationsergebnisse von EPIC und BGP beim Umschalten auf einen Ersatzpfad gemeinsam dargestellt. Dabei gilt es wieder den Einfluß des Simulationsmodells zu berücksichtigen.

Wie in Abschnitt 3.5.3.2 erklärt, spiegeln die “korrigierten” Werte das tatsächliche Verhalten wider. Auf die Konvergenzzeit von BGP und EPIC hat der MRAI nur eine verzögernde Wirkung. Allerdings konvergiert EPIC deutlich schneller, da beim Bekanntgeben des Ersatzpfads sofort alle ungültigen Pfade gelöscht werden. Somit zeigt sich an diesen Ergebnissen, daß das Pfadtesten



(a) (b)

Abbildung 3.18: Konvergenzzeit von EPIC, wenn ein Präfix nicht mehr (a) und wieder (b) erreichbar ist

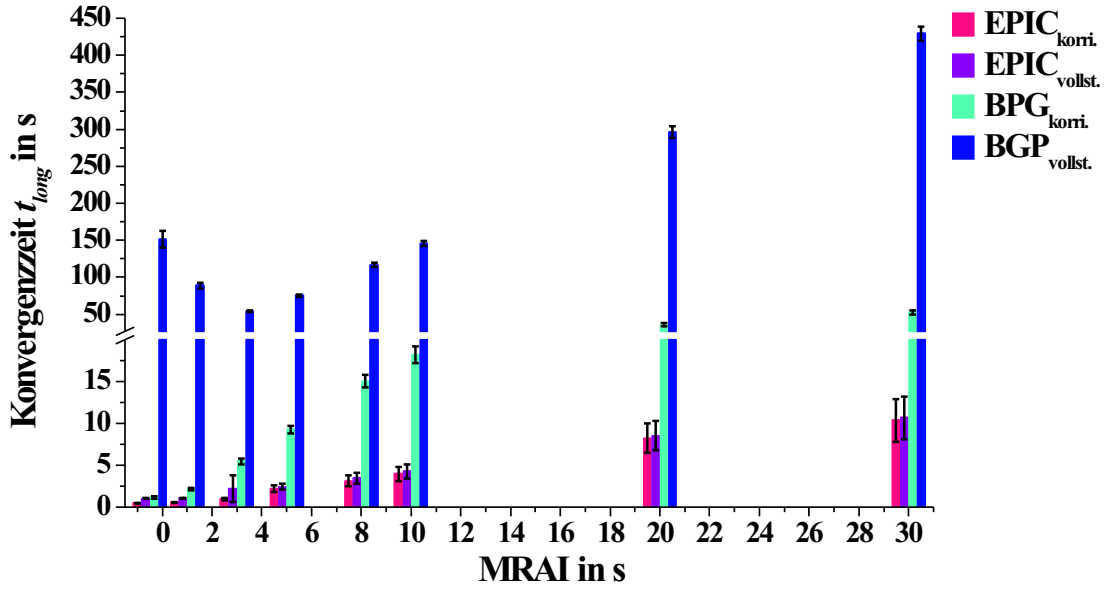
die Konvergenzzeit von BGP auch beim Umschalten auf einen Ersatzpfad verzögert. Wesentlich stärker zeigen dies auch die "vollständigen" Ergebnisse mit deutlich größeren Verbesserungen durch den Einsatz von EPIC.

Beim Umschalten auf einen besseren Pfad nach dessen Reparatur zeigt sich keinerlei Unterschied zwischen dem Konvergenzverhalten von BGP und EPIC. Die Ergebnisse der entsprechenden Simulationen sind in Abbildung 3.20 zu sehen.

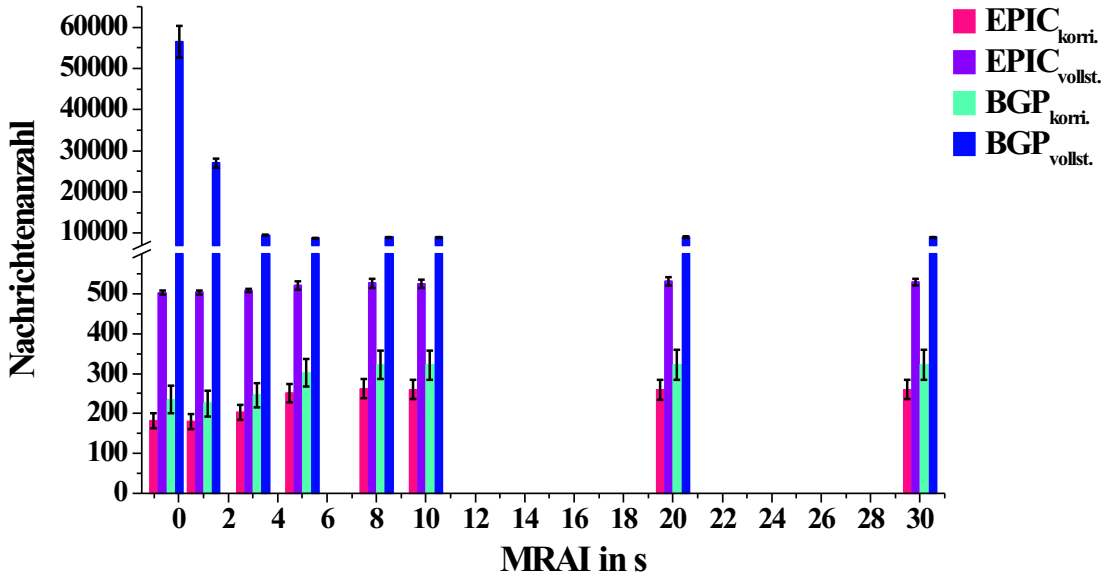
Die sich ergebenden Unterschiede zwischen den "korrigierten" und "vollständigen" Werten sind schon in Abschnitt 3.5.3.2 ausführlich behandelt worden und gelten für diese Simulationen unverändert. Abschließend kann auch hier festgestellt werden, daß der MRAI sich nur verzögernd auf die Konvergenzzeit von EPIC auswirkt.

3.5.5.2 Einordnung der Ergebnisse

Anhand der simulativen Untersuchung der Konvergenzzeit von EPIC konnte die Aussage, daß mittels des Einsatzes von EPIC das BGP-Pfadtesten vermieden werden kann, bestätigt werden. Desweiteren erzielte EPIC die besten Konvergenzzeiten bei allen vier Szenarien, wenn der MRAI komplett deaktiviert wurde. Da bei den Simulationen die interne Topologie nicht mit betrachtet wurde, muß man die Verweildauer einer BGP-UPDATE-Nachricht in einem BGP-Router und die Sendezeit zwischen dem Eingangs- und Ausgangs-Router zu der Konvergenzzeit hinzuaddieren. Damit entspricht die Konvergenzzeit von EPIC den Werten aus Abbildung 3.17 und erreicht die untere Grenze der Konvergenzzeit, wie sie in Abschnitt 3.4.5 abgeschätzt wurde.

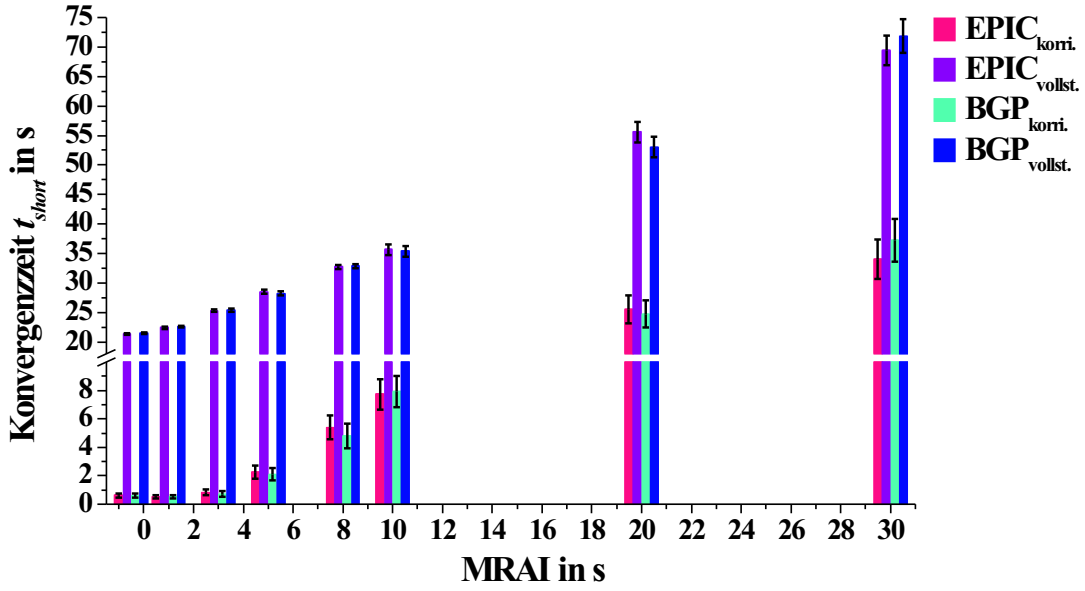


(a)

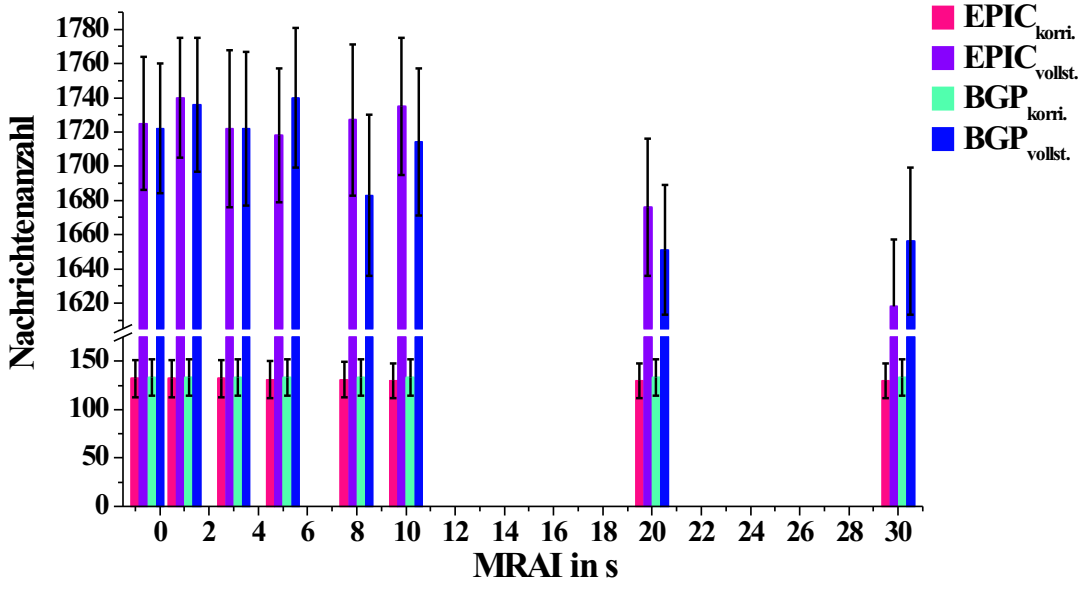


(b)

Abbildung 3.19: Konvergenzzeit (a) und Nachrichtenanzahl (b) von EPIC und BGP beim Umschalten auf einen Ersatzpfad



(a)



(b)

Abbildung 3.20: Konvergenzzeit (a) und Nachrichtenanzahl (b) von EPIC und BGP beim Umschalten auf einen besseren Pfad

3.6 BGP-Routing mit minimaler Konvergenzzeit

Wie in Abschnitt 3.3.7 erläutert wurde, ist eine minimale Konvergenzzeit von BGP aufgrund der Skalierungsanforderungen nur durch die Beschleunigung der Verteilung der Routing-Informationen möglich. Insofern gilt die in Abschnitt 3.4.5 ermittelte untere Grenze für die Konvergenzzeit.

Um diese untere Grenze zu erreichen, muß das BGP-Routing mittels EPIC erweitert werden. Damit wird sichergestellt, daß das BGP-Pfadtesten vermieden wird und somit der MRAI ausgeschaltet werden kann, wie es im letzten Abschnitt und in [**Sch06a**] belegt wurde. Diese Erweiterungen bilden die Grundlage für das BGP+-Routing. Der Name BGP+ soll verdeutlichen, daß es sich dabei weiterhin um ein BGP-Routing handelt, das um einige Mechanismen erweitert und die Konfiguration von BGP darauf abgestimmt wurde.

Zur Minimierung der Konvergenzzeit gehört auch eine schnelle Fehlererkennung, die durch den Einsatz der BFD, die in Abschnitt 3.3.1 schon erläutert wurde, realisiert werden kann. Somit ist die BFD auch ein wichtiger Bestandteil von BGP+. Deswegen spielt der Wert des Zeitglieds HoldTime nur noch eine untergeordnete Rolle beim BGP+-Routing.

Neben der BFD führen die Ansätze zur Nutzung redundanter Komponenten und Verbindungen (s. Abschnitt 3.3.3) sowie zur Vermeidung von sich ständig ändernden Pfaden durch das RFD+ (s. Abschnitt 3.3.2) auch zu einer Verbesserung des Konvergenzverhaltens. Deswegen werden diese Lösungen, so wie sie beschrieben wurden, in BGP+ übernommen.

Einen wichtigen, bisher unbetrachteten Aspekt stellt die Abwärtskompatibilität und die Einführungsstrategie von EPIC und damit auch von BGP+ dar. Durch die Erweiterung von BGP+ mit EPIC ist ein abgeändertes Format der BGP-UPDATE-Nachricht notwendig. Im Gegensatz zu den Beschreibungen in [PAM+05] reicht es nicht aus, ein neues optionales Attribut für den BGP-Pfadvektor zu definieren. Dieses Vorgehen hätte den Vorteil, daß man zum bestehenden BGP abwärtskompatibel bleiben könnte. Da aber die Information des Ursprungs der Topologieänderung auch im Withdrawn-Teil der BGP-UPDATE-Nachricht benötigt wird, reicht die Definition eines neuen optionalen Attributs nicht aus, weil dieses nur im Announcement-Teil der BGP-UPDATE-Nachricht verwendet werden kann. Es muß auch der Withdrawn-Teil der BGP-UPDATE-Nachricht erweitert werden, um einem Präfix, das gelöscht werden soll, den Ursprungs der Topologieänderung hinzufügen zu können. Die naheliegende Lösung ist die Definition von optionalen Attributen auch für den Withdrawn-Teil der BGP-UPDATE-Nachricht, wie es in Abbildung 3.21 dargestellt ist.

Mit dieser Erweiterung kann die Abwärtskompatibilität nicht gewährleistet werden. Da aber die Information über den Ursprung der Topologieänderung essentiell für das Erreichen der minimierten Konvergenzzeit ist, muß beim Aufbau der BGP-Verbindung geklärt werden, ob ein BGP-Router dieses erweiterte Format der BGP+-UPDATE-Nachricht unterstützt oder nicht. Dazu ist ein weiterer optionaler Parameter in der BGP-OPEN-Nachricht notwendig, deren Aufbau in Abbildung 3.22 dargestellt ist.

Nach [CS02] muß der neue Parameter die Form <Parametertyp, Länge des Parameters, Wert des Parameters> besitzen. Der Parametertyp muß bei der IETF festgelegt werden, damit dieser ein Byte lange Code eindeutig ist und nicht falsch interpretiert werden kann. Für die Länge

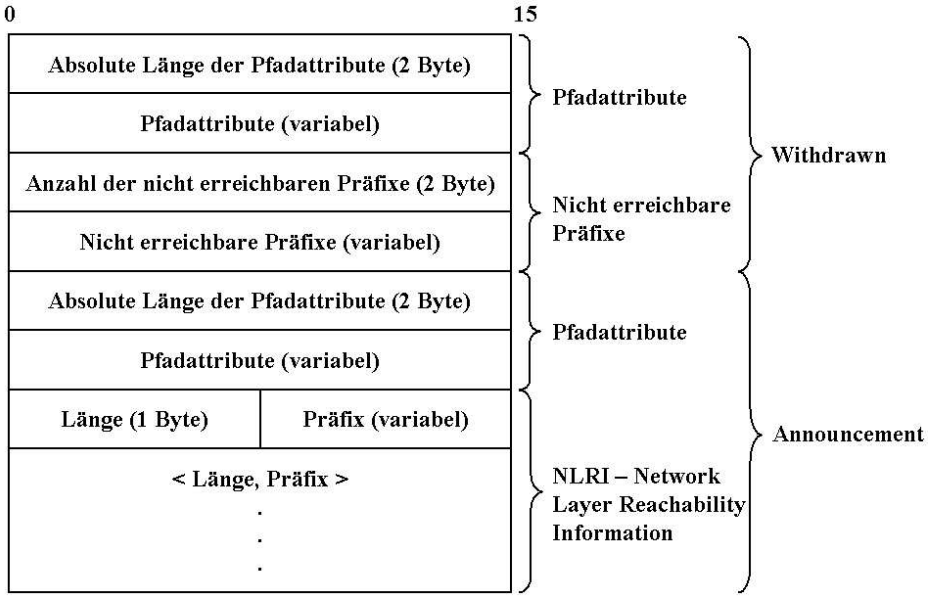


Abbildung 3.21: Format der BGP+-UPDATE-Nachricht

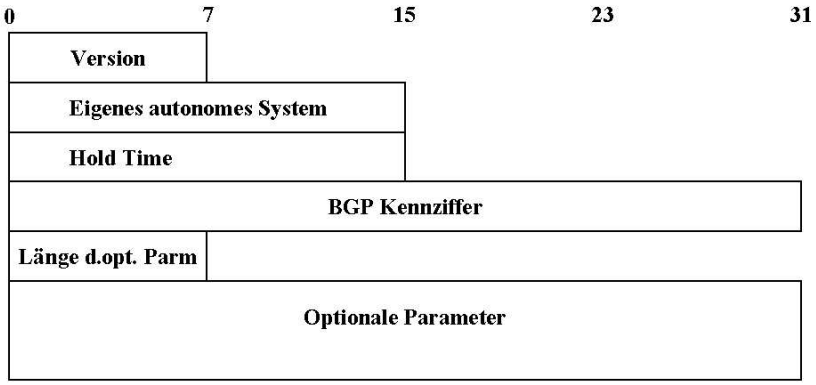


Abbildung 3.22: Format der BGP-OPEN-Nachricht nach [Hal97]

des Parameters reicht ein Bit aus, da nur mitgeteilt werden muß, ob das Format der erweiterten BGP+-UPDATE-Nachricht unterstützt wird ($\text{Wert} = 1$) oder nicht ($\text{Wert} = 0$). Der Vorteil dieses neuen optionalen Parameters ist, daß ein BGP-Router, der diesen Parameter nicht kennt, diesen einfach ignoriert und das Standard-BGP verwendet. Versteht ein Router BGP+ und damit das neue Attribut, so kann er mit dem Nachbar-Router mittels BGP+ die Routing-Informationen austauschen.

Somit wird sichergestellt, daß nur da BGP+ benutzt wird, wo es auch unterstützt wird. Allerdings kann BGP+ nicht seine volle Funktionalität entfalten, wenn andere autonome Systeme BGP+ nicht unterstützen. Aber mit steigender Anzahl an autonomen Systemen, die BGP+ einsetzen, kann die Konvergenzzeit wie beschrieben reduziert werden, da das Pfadtesten verringert bzw. vermieden wird.

Die hier vorgeschlagenen Erweiterungen ermöglichen eine sanfte Migration von BGP zu BGP+, wo beide Routing-Protokolle nebeneinander in verschiedenen autonomen Systemen des Internets eingesetzt werden können.

3.7 Zusammenfassung

In diesem Kapitel wurde das Konvergenzverhalten von BGP umfassend untersucht und auf die Eignung für den Datentransport mit garantierter Dienstqualität hin überprüft. Anhand der Messungen von Labovitz wurde klar, daß die Konvergenzzeit von BGP in einem inakzeptablen Bereich von mehreren Minuten liegt. Die Bewertung existierender Lösungsansätze hat ergeben, daß viele Mechanismen nur eingeschränkt, z.B. nur für bestimmte Fehlerbilder, in der Lage sind, die Konvergenzzeit zu reduzieren. Desweiteren konnte aus dieser Bewertung die Schlußfolgerung gezogen werden, daß nur die Beschleunigung der Verteilung der BGP-Routing-Informationen zu einer uneingeschränkten Verringerung der BGP-Konvergenzzeit führen kann.

Einen entscheidenden Einfluß auf die BGP-Konvergenzzeit hat das BGP-Zeitglied MRAI. Ohne dessen Nutzung sind heutige BGP-Router nur eingeschränkt in der Lage, die durch das BGP-Pfadtesten generierte Flut von BGP-UPDATE-Nachrichten zu verarbeiten. Anhand von Simulationsergebnissen konnte gezeigt werden, daß bei der BGP-Konvergenz, bei der das BGP-Pfadtesten auftritt, die optimale Konfiguration des MRAI zu einer Reduktion der Nachrichtenanzahl und der Konvergenzzeit führt. Kommt es aber bei der BGP-Reaktion auf eine Topologieänderung nicht zum BGP-Pfadtesten, dann verzögert der MRAI die BGP-Konvergenzzeit, ohne einen Vorteil aufweisen zu können.

Um die Konvergenzzeit von BGP zu minimieren, ist es notwendig, den MRAI zu deaktivieren. Anhand der durchgeführten Simulationen wurde gezeigt, daß dies nur erreicht werden kann, wenn das BGP-Pfadtesten vermieden wird. Diese Forderung läßt sich durch den Einsatz der BGP-Erweiterung EPIC realisieren, die nachweisbar das BGP-Pfadtesten verhindert. Dadurch ist es möglich, die BGP-Konvergenzzeit bis zur unteren Grenze zu reduzieren, die in dieser Arbeit theoretisch abgeschätzt wurde.

Aufgrund der Verwendung von realitätsnahen Internet-Topologien und der Berücksichtigung der internen Topologie eines autonomen Systems können diese Simulationsergebnisse auf das reale Internet übertragen werden. Das führt zu der Schlußfolgerung, daß der Einsatz von EPIC in Kombination mit deaktivierten MRAI eine minimale Konvergenzzeit im Bereich von einigen Sekunden ermöglicht.

Im letzten Teil dieses Kapitels wurden diese Erkenntnisse zu dem erweiterten Routing-Protokoll BGP+ zusammengefaßt. Darüber hinaus wurden die notwendigen Änderungen von BGP+ spezifiziert, damit es im heutigen Internet eingesetzt werden kann und dabei die problemlose Zusammenarbeit mit Standard-BGP garantiert wird.

Kapitel 4

Konzept für Garantierte Interdomain-Dienstqualität

In Kapitel 3 wurde das Interdomain-Routing mit BGP ausführlich analysiert und daraus das verbesserte Interdomain-Routing-Protokoll BGP+ abgeleitet. Basierend auf BGP+ wird in diesem Kapitel ein Konzept für die Gewährleistung der Dienstqualität zwischen mehreren autonomen Systemen entwickelt. Neben BGP+ ist dafür auch eine Interdomain-Ressourcenreservierung notwendig, für die weitere Protokolle und Mechanismen benötigt werden.

Ausgangspunkt für die Entwicklung dieses Konzeptes sind die Rahmenbedingungen, um zwischen mehreren autonomen Systemen die Dienstqualität sicherzustellen. Daran schließt sich die Untersuchung und die Bewertung verschiedener existierender Architekturen bzw. Protokolle für die Zusage der Interdomain-Dienstqualität bzw. zur Verwaltung der Interdomain-Ressourcen an. Da die bestehenden Ansätze höchstens nur in Teilaspekten die aufgestellten Rahmenbedingungen erfüllen, wird das Konzept für Garantierte Interdomain-Dienstqualität (KGID) entwickelt, das außer BGP+ auch noch einige Mechanismen der existierenden Ansätze weiterverwendet. Ein bisher unbetrachtetes Problem ist die Aufrechterhaltung der Zusage der Interdomain-Dienstqualität im Fall einer Interdomain-Routing-Änderung. Basierend auf der Abschätzung der Interdomain-Verkehrsverschiebungen aufgrund eines Interdomain-Verbindungsausfalls werden Dimensionierungsregeln erarbeitet, die dieses Problem lösen.

4.1 Rahmenbedingungen

Bei den folgenden Betrachtungen wird das heutige Internet als Ausgangsbasis für den Aufbau einer Infrastruktur für die Garantie der Interdomain-Dienstqualität angenommen. Im Gegensatz zur Gewährleistung der Dienstqualität innerhalb eines Netzes müssen mehrere Netzbetreiber zusammenarbeiten, um die Interdomain-Dienstqualität im Internet sicherzustellen.

Die bestehende Arbeitsweise des heutigen Internets wird im folgenden als Best Effort (BE) bezeichnet. Wenn ein autonomes System die Dienstqualität für den entsprechenden QoS-Verkehr

zusichern kann, wird dieses autonome System ein *Next Generation Network* (NGN) genannt. Es ist allerdings nicht ausreichend, daß ein autonomes System nur die Dienstqualität sicherstellen kann, um als NGN bezeichnet zu werden. Vielmehr stellt der Begriff "NGN" wesentlich höhere Anforderungen, wobei diese auch stark differieren können. Deswegen wurde von mehreren Organisationen definiert, was man unter einem NGN versteht. Einen Überblick über diese Aktivitäten wird in [LK05] gegeben.

Diese Arbeit verwendet den Begriff NGN, wie er in [IT04a, IT04b] standardisiert wurde. Dabei versteht man unter einem NGN ein paketbasiertes Netz, das die verschiedensten Kommunikationsdienste unterstützt. Dafür muß ein NGN unterschiedliche, breitbandige Transporttechnologien, die die Dienstqualität gewährleisten müssen, nutzen können. Desweiteren wird zwischen dem Datentransport und der Servicebereitstellung unterschieden und gefordert, daß ein Betreiber unabhängig vom Datentransport einen Service anbieten bzw. ein Netzbetreiber unabhängig von den Serviceangeboten den Datentransport sicherstellen muß. Außerdem muß ein NGN generell die Mobilität der Kunden unterstützen.

Neben den genannten Anforderungen wird in dieser Arbeit erwartet, daß ein NGN in der Lage ist, diese Funktionalität mit Ausnahme der Mobilität auch zwischen verschiedenen NGN zu ermöglichen. Was dafür notwendig ist, wird in diesem Kapitel erarbeitet und in das KGID integriert.

4.1.1 QoS-Inseln

Aufgrund der Größe und der Funktionsweise des Internets ist es nicht möglich, daß alle Netzbetreiber zu einem bestimmten Zeitpunkt ihre Netze auf die Unterstützung zur Garantie der Dienstqualität umstellen werden. Somit wird es für eine gewisse Zeitdauer ein Nebeneinander der heutigen sowie der neuen Technologie geben. Wie diese Koexistenz von BE- und NGN-Netzen aussehen kann, ist in Abbildung 4.1 beispielhaft dargestellt.

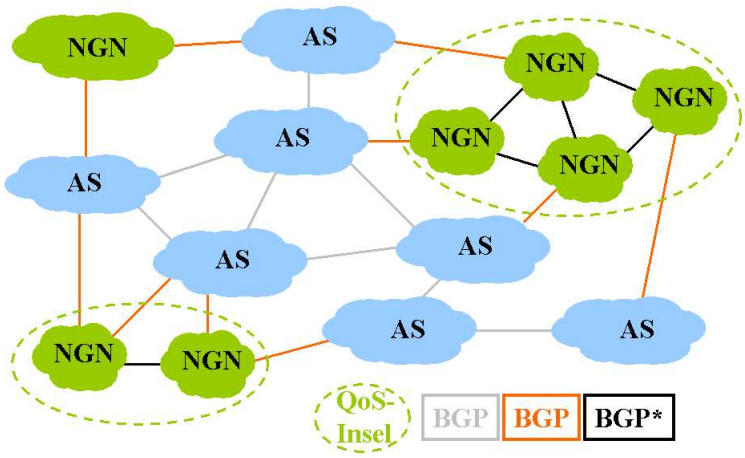


Abbildung 4.1: QoS-Inseln

Die Abbildung 4.1 zeigt auch, daß sich in unterschiedlichen Teilen des Internets NGN entwickeln können und diese Teile nicht zwangsweise miteinander verbunden sein müssen. Allerdings wird eine Interdomain-Kommunikation, die die Dienstqualität gewährleisten kann, nur möglich, wenn die NGN untereinander verbunden sind. Dies führt zu den hier als QoS-Inseln bezeichneten Gruppen von NGN. Diese entstehen aus einem einzelnen NGN, das sich mit der Zeit mit weiteren NGN verbindet, und somit vergrößert sich die QoS-Insel, bis sie am Schluß zu einem großen Netz zusammengewachsen ist.

Die QoS-Inseln haben gemeinsam, daß sie alle die Dienstqualität innerhalb eines und zu den benachbarten NGN zusagen können. Dafür ist die Einführung von neuen bzw. verbesserten Mechanismen notwendig. Das allmähliche Entstehen und Entwickeln der QoS-Inseln ermöglicht es, neue Verfahren einzuführen, da diese nicht zwangsweise kompatibel zu den BE-Netzen sein müssen. Es muß nur sichergestellt sein, daß die Interdomain-Kommunikation zwischen den NGN funktioniert und in allen QoS-Inseln die gleichen Mechanismen für die Interdomain-Kommunikation verwendet werden.

Besonders wichtig ist, daß die eingesetzten Mechanismen auch für den Fall funktionieren, daß das komplette Internet aus NGN besteht und somit eine einzige große QoS-Insel darstellt. Die daraus resultierenden Anforderungen sind schon in Abschnitt 3.1 näher betrachtet worden.

Ungeachtet der Anforderungen und Einschränkungen, die an ein NGN gestellt werden, muß der Netzbetreiber eines NGN die Freiheit besitzen zu entscheiden, welche Mechanismen und Protokolle er innerhalb seines NGN einsetzt, um die in den SLA zwischen den NGN festgeschriebene Dienstqualität und Fehlerrobustheit zu erfüllen.

4.1.2 Routing innerhalb der QoS-Inseln

Bei der Festlegung der Interdomain-Kommunikation zwischen den NGN darf eine wichtige Anforderung nicht außer Betracht gelassen werden. Die QoS-Inseln sind immer noch ein Bestandteil des Internets, und die Kunden der NGN wollen genauso wie die Kunden der autonomen Systeme alle Dienste nutzen, für die die BE-Datenübertragung ausreicht. Deswegen muß die Interdomain-Kommunikation der QoS-Inseln sicherstellen, daß zwischen den autonomen Systemen und den NGN weiterhin Daten ausgetauscht werden können. Dafür dürfen die autonomen Systeme keinen Unterschied feststellen, ob sie Daten mit einem anderen autonomen System oder einem NGN austauschen. Dies erfordert, daß die NGN an den Übergängen zu den autonomen Systemen BGP ohne jegliche Anpassungen einsetzen, was durch die roten Verbindungen in Abbildung 4.1 verdeutlicht wird.

Unabhängig davon erfordert das Routing zwischen den NGN einer QoS-Insel ein erweitertes BGP-Routing, um sicherzustellen, daß der QoS-Verkehr nicht die QoS-Insel verläßt, da die benachbarten autonomen Systeme die Dienstqualität des QoS-Verkehrs nicht gewährleisten können. Das erweiterte BGP-Routing der QoS-Inseln muß zwischen Pfaden unterscheiden, die die Dienstqualität garantieren, und denen, die nur BE-Verkehr transportieren können. Dieses erweiterte BGP-Routing wird in Abbildung 4.1 als BGP* bezeichnet.

Wie schon erwähnt, bieten die QoS-Inseln generell die Möglichkeit, Verbesserungen oder sogar neue Mechanismen innerhalb der QoS-Insel einzuführen, und damit stellt diese Anforderung kein Problem dar. Außerdem kann man diese Möglichkeit nutzen, um zusätzliche Erweiterungen, wie z.B. zur Reduzierung der BGP-Konvergenzzeit, einzuführen. Bevor die Umsetzung dieser Anforderungen beschrieben wird, folgt eine Analyse der bestehenden Lösungsansätze.

4.2 Bestehende Lösungsvorschläge

Den Stand der Wissenschaft repräsentiert hauptsächlich ein von der Europäischen Union geförderdtes Projekt, in dem ein Architekturansatz für die Unterstützung der Interdomain-Dienstqualität erarbeitet wurde. Darüber hinaus existieren noch auf der Architektur dieses Projektes aufbauende Ansätze sowie verschiedene Konzepte für die Interdomain-Ressourcenreservierung. Auf alle diese Lösungsvorschläge wird im folgenden kurz eingegangen.

4.2.1 Das MESCAL-Konzept

Das von der Europäischen Union geförderte Projekt lief unter der Bezeichnung “Management of End-to-end Quality of Service Across the Internet at Large” (MESCAL) [HFP+05] und baut auf den Ergebnissen des Projekts “TEQUILLA” [Myk03] auf. Bei TEQUILLA lag das Hauptaugenmerk auf der Unterstützung der Garantie der Dienstqualität innerhalb eines Netzes, und MESCAL erweitert dies auf die Interdomain-Kommunikation.

Ausgangspunkt von MESCAL ist die Ressourcenreservierung für Anwendungen, die eine Zusage der Dienstqualität zwischen verschiedenen autonomen Systemen benötigen. Dabei wird in MESCAL das Konzept der Klasseneinteilung der DiffServ-Architektur (s. Abschnitt 2.2.2.2) angewendet. Allerdings werden die QoS-Klassen für die Interdomain-Kommunikation verwendet und als *meta-QoS-classes* [LBM+04] bezeichnet.

Darüber hinaus setzt das MESCAL-Konzept voraus, daß die Dienstqualität auch innerhalb eines autonomen Systems mit dem DiffServ-Ansatz sichergestellt wird. Die in einem autonomen System verwendeten QoS-Klassen werden als *local-QoS-classes* (l-QC) bezeichnet. Ob ein autonomes System die Dienstqualität garantieren kann, ist Bestandteil des SLA zwischen zwei benachbarten autonomen Systemen, wobei die *Service Level Specification* (SLS), die ein Teil des SLA ist, die technischen Anforderungen der Dienste festlegt, wie z.B. Bandbreite oder Verzögerung. Diese SLS werden noch in zwei unterschiedliche Typen unterteilt. Die Dienstqualität zwischen dem Kunden und einem autonomen System wird in dem sogenannten *Customer SLS* (cSLS) beschrieben, und die QoS-Anforderungen zwischen zwei autonomen Systemen werden im *Peer SLS* (pSLS) festgelegt.

Basierend auf den local-QoS-classes wird die Dienstqualität eines Interdomain-Pfads bestimmt, was anhand eines Beispiels aus [MBA+04] in Abbildung 4.2 dargestellt ist.

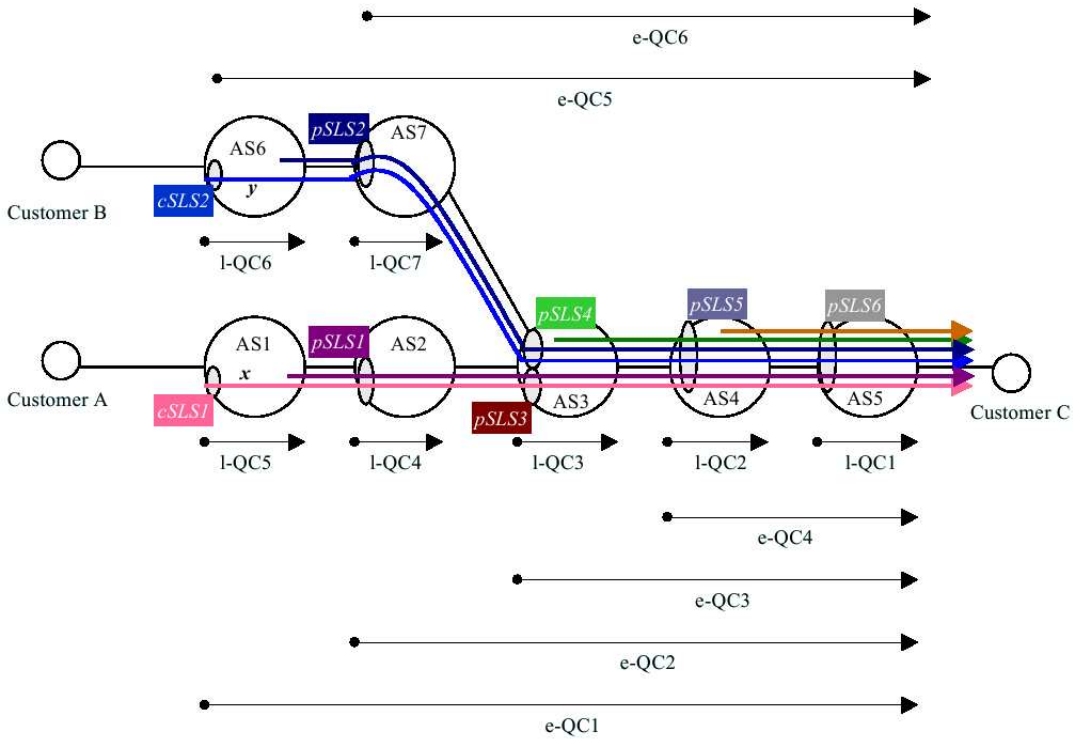


Abbildung 4.2: Auswahl der QoS-Klassen eines Pfads beim MESCAL-Konzept [MBA+04]

In diesem Beispiel sieht man, wie man anhand der local-QoS-classes die für den Interdomain-Pfad gültigen QoS-Klassen – die *extended-QoS-classes* (e-QC) – ermitteln kann. Die extended-QoS-classes stellen dabei die Gesamtheit aller local-QoS-classes der durchlaufenen autonomen Systeme dar. Ist der betrachtete Parameter der QoS-Klasse die Verzögerung, so werden die Verzögerungen der einzelnen local-QoS-classes zu der Verzögerung der extended-QoS-classes des Pfads zusammengezählt. Wie die sonstigen Kombinationen gebildet werden, ist ausführlich in [MBC+05] beschrieben.

Diese Form der Ressourcenreservierung stellt auch besondere Anforderungen an das Interdomain-Routing. Deswegen wurde das QoS enhanced BGP (q-BGP) [Bou05] entworfen, welches zu den Erreichbarkeitsinformationen zusätzlich noch Parameter für die Bestimmung der Dienstqualität eines Pfads verteilt. Auf die genaue Funktionsweise wird im nächsten Abschnitt näher eingegangen.

Daneben wird für das Interdomain-Verkehrsmanagement vorausgesetzt, daß q-BGP mehrere Pfade zu einem Präfix verwalten kann. Die darauf aufbauende Verkehrsverteilung stellt einen großen Schwerpunkt des MESCAL-Konzepts dar, und die Ergebnisse sind in [MBC+05, HHW+05, HPH+04] beschrieben. Um mehrere Pfade zu einem Präfix beim Interdomain-Routing benutzen zu können, setzt man in MESCAL auf den Ansatz aus [WRC06] auf. Außerdem geht man davon aus, mit diesem Konzept mehrere Pfade zu einem Präfix zu erhalten, um einen dieser Pfade als Ersatzpfad im Fehlerfall nutzen zu können.

4.2.2 QoS Enhanced BGP

Um die charakteristischen Parameter eines Pfads, die für die Ressourcenreservierung von Interesse sind, zu verteilen, wurde im Rahmen des MESCAL-Konzepts das q-BGP entwickelt. Der wesentliche Unterschied zu BGP ist ein zusätzliches Attribut in der BGP-UPDATE-Nachricht, was in [CJ03] als IETF Internet Draft beschrieben ist. Dieses Attribut kann die folgenden Parameter enthalten:

- Reserviert
- Reservierte Bandbreite
- Verfügbare Bandbreite
- Verlustwahrscheinlichkeit
- minimale Verzögerung in eine Richtung
- maximale Verzögerung in eine Richtung
- mittlere Verzögerung in eine Richtung

Somit können mittels der BGP-UPDATE-Nachrichten die für die Ressourcenreservierung interessanten Parameter zur Verfügung gestellt werden. Desweiteren werden diese Parameter auch bei der Auswahl des besten Pfads berücksichtigt und ermöglichen damit ein Interdomain-Verkehrsmanagement anhand der QoS-Parameter eines Pfads.

Für die Suche der besten Pfade von q-BGP werden zwei verschiedene Gruppen von Algorithmen vorgeschlagen. Die Gruppe 1 erfüllt die Anforderung aus Abschnitt 4.1.2, daß zwischen den Pfaden unterschieden werden muß, wo die Dienstqualität zugesagt werden kann oder nicht. Im Fall von Gruppe 2 werden die zusätzlichen QoS-Parameter bei der Bestimmung der besten Pfade berücksichtigt. Bei der geänderten Suche der besten Pfade für die Gruppe 1 werden nach [MBC+05] die folgenden Übereinstimmungen überprüft:

1. Zu welchem Präfix gehört die empfangene Routing-Information?
2. Für welche meta-QoS-class ist die BGP-UPDATE-Nachricht bestimmt?
3. Erfüllt die BGP-UPDATE-Nachricht die lokalen Policies?
4. Anwenden der Standardsuche der besten Pfade von BGP (s. Abschnitt 2.1.2.2)

Betrachtet man die Suche der besten Pfade der Gruppe 2 [MBC+05], so wird der dritte Punkt bei der Bestimmung der besten Pfade von Gruppe 1 durch die Berücksichtigung der QoS-Parameter des entsprechenden Pfads ersetzt. Wie die verschiedenen Parameter gewichtet werden, wird in [MBC+05] ausführlich diskutiert.

Wie man an der Beschreibung der Suche der besten Pfade von q-BGP schon ablesen kann, werden dabei die Routing-Informationen nach der meta-QoS-class sortiert. Das hat zur Folge, daß für jede meta-QoS-class unterschiedliche Pfade benutzt werden können. Daraus resultiert, daß pro meta-QoS-class eine eigene BGP-Routing-Tabelle benötigt wird.

4.2.3 Auf q-BGP aufbauende Ansätze

Das im Rahmen des MESCAL-Projektes entwickelte q-BGP wird auch in weiteren Ansätzen verwendet, die im folgenden kurz vorgestellt werden.

4.2.3.1 Overlay-Netze

In [YFMB+04] wird eine Interdomain-Overlay-Architektur vorgestellt, die eine Kommunikation zwischen zwei nicht benachbarten autonomen Systemen unter Gewährleistung der Dienstqualität ermöglicht. Pro autonomes System wird dabei mindestens ein Overlay-Knoten, der vollen Zugriff auf die BGP-Routing-Tabellen benötigt, installiert. Zwischen den Overlay-Knoten verschiedener autonomer Systeme wird, wie im normalen Internet, ein SLA ausgehandelt. Um die Dienstqualität zusichern zu können, müssen die an der Kommunikation beteiligten autonomen Systeme ebenfalls diese Dienstqualität gewährleisten. Um dafür die verfügbaren Ressourcen auf den benutzten Pfaden zu erfahren, wird auf q-BGP aufgesetzt.

Der Vorteil der Overlay-Knoten besteht in der Möglichkeit, eine erweiterte Funktionalität einzuführen, ohne diese bei allen betroffenen autonomen Systemen ebenfalls voraussetzen zu müssen. In [YFMB+04] wird dies benutzt, um die Interdomain-Dienstqualität zu gewährleisten und um zwei verschiedene Pfade zu dem jeweiligen Partner-AS zu verwalten. Wie mit Simulation gezeigt wird, kann damit die Fehlerreaktion von BGP beschleunigt werden. Die Ergebnisse zeigen eine Verbesserung um 75s, wobei die Standardreaktion von BGP 80s benötigt. Allerdings sind die Angaben in [YFMB+04] zu dem Simulationsmodell nicht ausreichend, um diese Ergebnisse genauer zu bewerten. Aufgrund einer separaten Fehlererkennung der Overlay-Knoten und der Existenz eines zweiten Pfads erscheinen die Ergebnisse unter den getätigten Voraussetzungen plausibel.

Desweiteren unterliegt diese Architektur der Einschränkung, daß der Ansatz nur für mehrfach angebundene Stub-AS entwickelt worden ist. Nach Angaben in [YFMB+04] stellen diese 80% aller autonomen Systeme dar, so daß der Ansatz für die restlichen 20% nicht funktioniert.

Aufbauend auf dem Ansatz aus [YFMB+04] wird in [YFMB+05] ein selbstregulierender Mechanismus zur Netzoptimierung vorgestellt. Dabei wird je nach Lastsituation Verkehr zwischen den beiden Pfaden der jeweiligen Overlay-Knoten so verteilt, daß keiner dieser beiden Pfade überlastet wird.

4.2.3.2 ENTHRONE-Ansatz

Ein weiterer Ansatz, der auf q-BGP aufbaut, ist der ENTHRONE-Ansatz [Ent06]. Hierbei wird nicht nur q-BGP benutzt, sondern es wird das komplette Interdomain-Routing des MESCAL-Konzepts verwendet. Hauptmotivation ist der andere Fokus des ENTHRONE-Projekts, der auf die Übertragung von Audio- und Videodaten über mehrere Domänen und der Zusicherung der benötigten Ressourcen gerichtet ist. Damit wird der Schwerpunkt eher auf die Anwendungsebene gelegt, um die Interaktion zwischen Inhalteanbieter, Netzbetreiber, Diensteanbieter und Inhaltekonsument zu ermöglichen.

Für die Ressourcenreservierung wird ein ähnlicher Ansatz wie der des MESCAL-Konzepts verwendet. Die QoS-Klasse der Ende-zu-Ende-Übertragung wird ebenfalls kaskadenartig aus den lokalen QoS-Klassen der durchlaufenen autonomen Systeme zusammengesetzt. Allerdings berechnet dies nicht ein autonomes System für alle seine Pfade zu den verschiedenen Zielen, sondern der Diensteanbieter sucht den passenden Pfad zwischen dem Inhalteanbieter und dem Inhaltekonsumenten und bestimmt die Dienstqualität durch die Abfrage der am Datentransport beteiligten autonomen Systeme. Basierend auf diesen Informationen kann er die unterstützte Dienstqualität des Ende-zu-Ende Pfads ermitteln.

4.2.4 Interdomain-Ressourcenreservierung mit BGRP und SICAP

Eine Möglichkeit, um entlang eines Interdomain-Pfads Ressourcen zu reservieren, stellt das *Border Gateway Reservation Protocol* (BGRP) [PHS00] dar. Um nicht für jede einzelne Anwendung die belegten Ressourcen zu verwalten, versucht BGRP die angeforderten Ressourcen zu aggregieren.

Dazu verschickt der Sender, der Ressourcen für eine Anwendung reservieren will, eine *PROBE*-Nachricht zum Empfänger und testet, ob genügend Ressourcen verfügbar sind. Ist dem so, werden die benötigten Ressourcen mittels der *GRAFT*-Nachricht auf dem Rückweg belegt, was anhand der zwei Reservierungen A und B in Abbildung 4.3 aus [SGV03] dargestellt ist.

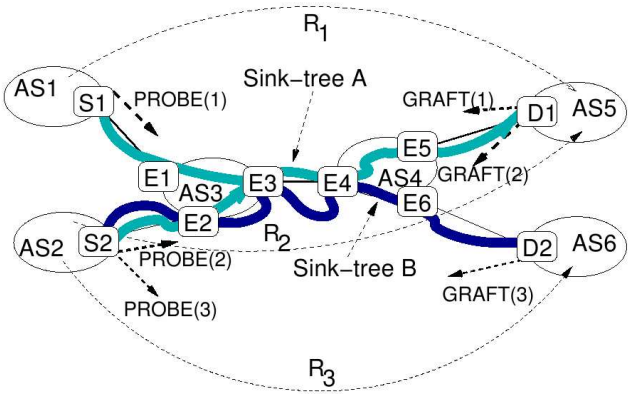


Abbildung 4.3: Ressourcenreservierung mit BGRP [SGV03]

Wird während des Überprüfens eine existierende Reservierung zu demselben Ziel gefunden, so wird diese benutzt und die schon belegten Ressourcen der neuen Anforderung angepaßt. Damit müssen nicht in jedem Knoten die entsprechenden Ressourcen für jede Anwendung verwaltet werden. Sondern es reicht aus, nur die aggregierten Verkehrsflüsse zu überwachen und so die Ressourcenreservierung zu vereinfachen.

Aufgrund dieser Vorgehensweise bildet sich ein Baum (*sink-tree*) von reservierten Ressourcen aus, wobei die Sender die Äste und der Empfänger die Wurzel des Baumes darstellen. Je mehr Sender Ressourcen zu einem Empfänger belegen, um so stärker ist die Aggregation der zu verwaltenden Reservierungsinformationen.

Wie in Abbildung 4.3 zu sehen ist, benutzen die Reservierungen A und B die gleiche Verbindung zwischen AS3 und AS4. Da diese beiden Reservierungen nicht das gleiche Ziel besitzen, kann BGRP diese nicht zusammenfassen. Um dies zu ermöglichen, wurde das *Shared-segment Interdomain Control Aggregation Protocol* (SICAP) [SGV03] entwickelt. Damit läßt sich der Aggregierungsgrad der Reservierungen deutlich steigern, was den Verwaltungsaufwand pro Knoten reduziert.

Die Anzahl der Signalisierungsnachrichten ändert sich jedoch damit kaum. Dies wird erst möglich, wenn eine gewisse Überdimensionierung der reservierten Ressourcen vorgenommen wird. Somit müssen nicht für jede neue Anfrage entlang des kompletten Pfads die vorhandenen Ressourcen überprüft werden. Sondern sobald man ein Aggregat mit genügend freien Ressourcen findet, kann das Überprüfen beendet und auf dem bis dahin passierten Teilpfad die entsprechenden Ressourcen belegt werden. Wieviel extra Ressourcen dafür notwendig sind, um den Aggregationsgrad zu optimieren, wird detailliert in [Sof03] untersucht.

4.2.5 Protokollfamilie der NSIS-Arbeitsgruppe

Die am weitesten fortgeschrittenen Arbeiten für die Definition einer Interdomain-Ressourcenreservierung sind Teil der Protokollfamilie, die in der Arbeitsgruppe *Next Steps in Signaling* (NSIS) [IET06] der IETF standardisiert wird. Der prinzipielle Aufbau dieser Protokollfamilie ist in [HKL+05] beschrieben und in Abbildung 4.4 aus [FSB+05] schematisch dargestellt.

Grundsätzlich besteht der Rahmen in Abbildung 4.4 aus zwei Schichten, dem *NSIS Transport Layer Protocol* (NTLP) und dem *NSIS Signaling Layer Protoco*l (NSLP). Die untere Schicht NTLP [SH06] wird hauptsächlich durch das *General Internet Signaling Transport* (GIST) Protokoll repräsentiert, das auf der TCP/IP Protokollfamilie aufsetzt.

GIST dient hauptsächlich dem Transport der Signalisierungsnachrichten der höheren Schicht, dem NSLP. Dafür stellt es eine einheitliche Schnittstelle zur Verfügung und erlaubt somit den Einsatz von verschiedenen Signalisierungsprotokollen, wie z.B. für die Ressourcenreservierung. Dieses Protokoll ist in [MKM06] beschrieben und dessen prinzipielle Funktionsweise zeigt Abbildung 4.5.

Dabei wird mittels GIST eine Kommunikation zwischen den beteiligten Knoten (*Flow End Point* und *Network Element* (NE)) hergestellt. Diese muß nicht zwischen benachbarten Kno-

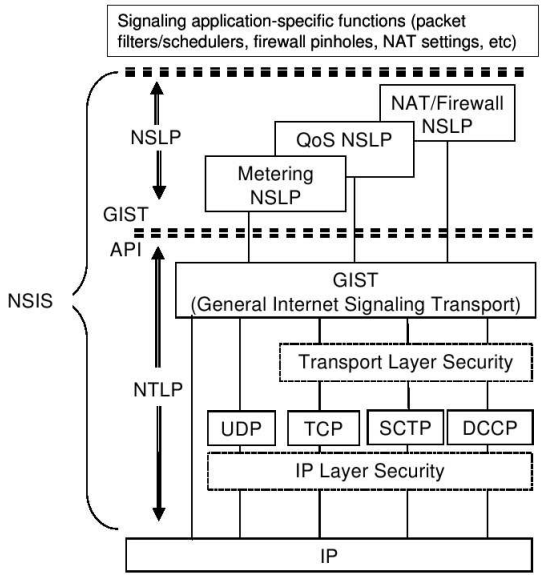


Abbildung 4.4: NSIS Protokollfamilie aus [FSB+05]

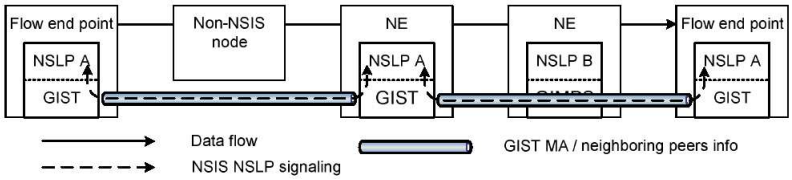


Abbildung 4.5: NSIS Ende-zu-Ende Kommunikation

ten erfolgen, sondern kann auch über Knoten laufen, die an dieser Signalisierung nicht beteiligt sind. Für die Knoten ist dies ganz normaler Datenverkehr. Speziell für die Interdomain-Ressourcenreservierung hat dies den Vorteil, daß nur der Eingangs- und der Ausgangs-Router eines autonomen Systems die Signalisierungsnachrichten verarbeitet und die restlichen Router dieses autonomen Systems unbeteiligt sind. Damit besteht die Möglichkeit, daß der Netzbetreiber eines autonomen Systems entscheiden kann, welche Mechanismen zur Ressourcenreservierung innerhalb des autonomen Systems eingesetzt werden.

4.2.6 Bewertung der existierenden Ansätze

In den letzten Abschnitten wurden die verschiedenen Ansätze für die Gewährleistung der Interdomain-Dienstqualität und für die Ressourcenverwaltung kurz vorgestellt. Dabei ermöglicht das MESCAL-Konzept die Interdomain-Ressourcenreservierung basierend auf den meta-QoS-classes zwischen den einzelnen autonomen Systemen. Der Ansatz, global eindeutige QoS-Klassen einzusetzen, ist die einzige Möglichkeit der Differenzierung der Interdomain-Dienstqualität, wie in Abschnitt 4.3.1 noch ausführlicher diskutiert wird.

Desweiteren werden für die Interdomain-Ressourcenreservierungen des MESCAL-Konzepts die Kenntnisse über die verfügbaren Ressourcen benötigt. Für die Übermittlung dieser Angaben wird auf das erweiterte BGP-Routing-Protokoll q-BGP gesetzt. Diese Vorgehensweise birgt das Risiko, daß die Aktualität der QoS-Parameter nicht sichergestellt werden kann, da die QoS-Parameter nur mit den BGP-UPDATE-Nachrichten bei der Änderung eines Pfads übermittelt werden. Damit erhält man keine periodische Aktualisierung dieser QoS-Parameter und läuft Gefahr, daß bei der Suche der besten Pfade der Gruppe 2 Pfade ermittelt werden, die die Dienstqualität aufgrund veralteter Informationen über die verfügbaren Ressourcen nicht sicherstellen können.

Nicht nur, daß bei der Bestimmung der besten Pfade der Gruppe 2 die Policies nicht mehr berücksichtigt werden, sondern diese Entscheidung ist auch von den QoS-Parametern der Pfade abhängig. In [FH99] wird gezeigt, daß dieses Routing zu Instabilitäten führen kann, wenn sich die QoS-Parameter zu schnell ändern. Dies wird auch in [MBC+05] als Problem erkannt, aber es wird keine Lösung dafür angeboten.

Betrachtet man hingegen die Suche der besten Pfade der Gruppe 1, erhält man eine Lösung, die die Anforderungen aus Abschnitt 4.1.2 erfüllt. Allerdings würde dafür auch eine binäre Unterscheidung zwischen einem Pfad, auf dem die Dienstqualität zugesagt werden kann, und einem BE-Pfad vollkommen ausreichen.

Außerdem wird bei dem MESCAL-Konzept angenommen, daß q-BGP eine eigene Routing-Tabelle für jede meta-QoS-class verwaltet und zu jedem Ziel mindestens zwei Pfade zur Verfügung stellt. Dies hat zur Konsequenz, daß q-BGP ein Mehrfaches an Speicher und Rechenleistung benötigt, um diese Anforderungen zu erfüllen. Der Speicherbedarf von q-BGP erhöht sich im Vergleich zu BGP um das Produkt der Anzahl der Pfade pro Präfix und der Anzahl der meta-QoS-classes. Geht man von fünf meta-QoS-classes und zwei Pfaden pro Präfix aus, dann vergrößert sich der Speicherbedarf von q-BGP um den Faktor zehn. Bei der Anzahl der q-BGP Nachrichten entspricht die Vergrößerung nur der Anzahl der meta-QoS-classes, da für die mehrfachen Pfade keine extra Nachrichten benötigt werden.

Damit wird deutlich, welche hohen Anforderungen dieser Ansatz an die Leistungsfähigkeit der q-BGP-Router stellt. Daraus resultierend muß man die Skalierfähigkeit von q-BGP aufgrund der Verwaltung von mehreren Pfaden pro Präfix und pro meta-QoS-class in Frage stellen.

Außerdem wurde in dieser Arbeit schon gezeigt, daß das BGP-Routing mit verschiedenen Pfaden zu einem Präfix nicht zwangsweise eine Verbesserung der BGP-Fehlerreaktion darstellt (s. Abschnitt 3.3.7).

Zusätzlich weist dieser Ansatz noch einige weitere ungeklärte Schwachstellen auf. Für die Unterscheidung der einzelnen Pfade zu einem Präfix wird in [WRC06] ein neues optionales Attribut eingeführt, womit die Abwärtskompatibilität zu BGP erhalten bleibt. Allerdings können im Withdrawn-Teil der BGP-UPDATE-Nachricht keine Attribute gesetzt werden. Damit stellt dies bei Ausfall eines der beiden Pfade ein Problem dar, da ein q-BGP-Router nicht entscheiden kann, welcher der beiden Pfade betroffen ist.

Eine weitere Schwierigkeit tritt aufgrund des zielbasierten Routing eines IP-Routers auf. Dieser kann nur anhand der Zieladresse eines IP-Paketes entscheiden, wohin es geschickt werden soll.

Somit besteht die Gefahr bei mehreren Pfaden mit einem Ziel, daß nicht nachvollziehbar ist, über welchen Pfad ein Paket weitergeleitet wird. Das wirkt sich nachteilig auf die Interdomain-Verkehrsverteilung aus, da nicht exakt festgelegt werden kann, welcher Verkehr welchen Pfad benutzt. Eine mögliche Lösung wäre der Einsatz von MPLS, wo anhand der benutzten MPLS-Pfade die Verkehrsverteilung nachvollziehbar ist. Aber die Beschränkung auf MPLS als Intradomain-Routing-Protokoll wurde in keiner der Beschreibungen des MESCAL-Konzepts erwähnt.

Zusammenfassend muß festgestellt werden, daß das q-BGP einige unübersehbare Mängel aufweist, für die keine Lösungen existieren, und allein das Nichtbeachten der Policies kann ausreichen, daß das gesamte Konzept von den Netzbetreibern nicht angenommen wird. Desweiteren gibt es mit dem Bereitstellen von mehreren Pfaden zu einem Präfix noch etliche ungeklärte Punkte, die den Einsatz von q-BGP noch weiter erschweren. Die Konsequenz aus der Gesamtheit der Schwachstellen führt dazu, daß das MESCAL-Konzept in dieser Arbeit nicht weiterverfolgt wird.

Der Ansatz des ENTHRONE-Projekts bietet eine weitere Lösung für die Interdomain-Ressourcenverwaltung an. Dieser beruht ebenfalls auf dem Interdomain-Routing mit q-BGP, wie es für die MESCAL-Architektur entwickelt wurde. Nicht nur, daß dieser Ansatz damit die gesamten Nachteile des MESCAL-Konzepts übernimmt, sondern das ENTHRONE-Konzept hat einen sehr speziellen Fokus auf die Interdomain-Übertragung von Audio- und Videodaten, was einer Berücksichtigung in dieser Arbeit entgegensteht.

Das Konzept der Overlay-Netze ist ein interessanter Ansatz, der ebenfalls für die Realisierung der Zusage der Interdomain-Dienstqualität genutzt werden kann. Allerdings ist es in den QoS-Inseln nicht unbedingt notwendig, eine weitere Schicht einzuführen, da es innerhalb der QoS-Inseln möglich ist, neue Mechanismen hinzufügen. Auch die Overlay-Netze sind an dieselben Randbedingungen wie die Skalierbarkeit gebunden. Somit ergibt sich kein Vorteil durch die Nutzung von Overlay-Netzen. Im Gegenteil, aufgrund der zusätzlichen Schicht erhöht sich eher die Komplexität. Außerdem setzen die Overlay-Ansätze aus Abschnitt 3.3.6 auch auf q-BGP auf, was, wie schon erwähnt, deutliche Nachteile mit sich bringt. Deswegen wird das Konzept der Overlay-Netze auch nicht weiter in die folgenden Überlegungen einbezogen.

BGRP und SICAP stellen zwei einfache Lösungen für die Interdomain-Ressourcenreservierung dar, die sich nur durch ihre Aggregationsfähigkeit unterscheiden. Beide setzen auf dem Standard-BGP-Routing auf und bieten ausschließlich Mechanismen, um zwischen verschiedenen autonomen Systemen Ressourcen zu belegen.

Die NSIS-Protokollfamilie stellt eine generelle Plattform für die Signalisierung innerhalb eines Netzes und zwischen verschiedenen autonomen Systemen zur Verfügung. Die Standardisierung bei der IETF stellt in Aussicht, daß diese Protokolle implementiert werden und somit in realen Netzen zum Einsatz kommen können.

Die untere Schicht, die als Schnittstelle zu den Transportprotokollen von IP fungiert, ist mittlerweile fertig definiert und das QoS-NSLP für die Interdomain-Ressourcenreservierung befindet sich im Standardisierungsprozeß. Damit ist dieser Ansatz am weitesten fortgeschritten und empfiehlt sich für die Verwendung in den QoS-Inseln. Außerdem bietet die NSIS-Lösung auch noch weitere Fähigkeiten, wie z.B. die Funktionalitäten für *Authentication*, *Authorization* und *Accounting*.

4.3 Entwurf des KGID

Die bestehenden Ansätze für die Gewährleistung der Interdomain-Dienstqualität und für die Ressourcenverwaltung ermöglichen zumindest die Ressourcenreservierung auf bestehenden Pfaden. Allerdings weisen die Routing-Ansätze einige zum Teil gravierende Mängel auf. Deswegen wird im folgenden Abschnitt das KGID entwickelt, das die in Abschnitt 4.1 formulierten Anforderungen einer QoS-Insel erfüllt und einige bestehende Ansätze aus dem letzten Abschnitt miteinbezieht.

4.3.1 Gewährleistung der Interdomain-Dienstqualität

Wegen der Skalierbarkeitsanforderungen der QoS-Inseln kann die Dienstqualität zwischen den NGN nur in verschiedene QoS-Klassen wie bei der DiffServ-Architektur (s. Abschnitt 2.2.2.2) eingeteilt werden. Allerdings müssen die Interdomain-QoS-Klassen aller QoS-Inseln die gleichen sein, wie es auch bei dem MESCAL-Konzept der Fall ist.

Werden die Interdomain-QoS-Klassen nicht festgeschrieben und jedes NGN kann diese mit seinen Nachbarn aushandeln, wäre es möglich, daß zwischen zwei NGN für den gesamten QoS-Verkehr nur eine QoS-Klasse verwendet wird. Würde jetzt QoS-Verkehr über diese Verbindung geleitet, könnte das NGN, das den QoS-Verkehr empfängt, diesen QoS-Verkehr beim Verlassen des NGN nicht mehr in verschiedene Interdomain-QoS-Klassen aufteilen, da die dafür notwendigen Informationen nicht mehr vorliegen. Dies hätte zur Konsequenz, daß auf dem weiteren Pfad der QoS-Verkehr nur einer QoS-Klasse zugeordnet wird und so keine Differenzierung des Verkehrs mehr stattfindet. Aus diesem Grund sind einheitliche Interdomain-QoS-Klassen in allen QoS-Inseln unumgänglich. Das bedeutet aber nicht, daß ein Netzbetreiber in seinem autonomen System DiffServ für die Sicherstellung der Dienstqualität einsetzen muß. Desweiteren wird mit der Definition der Interdomain-QoS-Klassen nicht vorgeschrieben, daß diese auch innerhalb eines autonomen Systems verwendet werden.

Bei der Definition der Interdomain-QoS-Klassen muß auch eine eindeutige Kennzeichnung in Form des DSCP [NBB+98] festgelegt werden. Anhand des DSCP kann jeder Eingangs-Router erkennen, welcher QoS-Klasse der Verkehr zuzuordnen ist und wie dieser Verkehr verarbeitet werden muß. Dafür ist es auch notwendig, daß der Ausgangs-Router den DSCP korrekt beim Verlassen des NGN setzt. Handelt es sich um QoS-Verkehr, muß die richtige Interdomain-QoS-Klasse zugewiesen werden. Bei der Weiterleitung von BE-Verkehr gilt es zu unterscheiden, ob dieser Verkehr zu einem NGN oder zu einem normalen autonomen System weitergeleitet werden soll. Wird der Verkehr an ein NGN geschickt, muß der DSCP für BE-Verkehr gesetzt und im anderen Fall sollte der DSCP gelöscht werden, um ein Fehlverhalten auszuschließen.

Für die exklusive Vergabe der für die Garantie der Interdomain-Dienstqualität benötigten Ressourcen wird in allen QoS-Inseln ein Protokoll für die Interdomain-Ressourcenreservierung benötigt. Dieses muß für alle NGN dasselbe sein, damit auf allen Pfaden der QoS-Inseln Ressourcen reserviert werden können. Aufgrund der Skalierbarkeitsanforderungen ist es unmöglich,

Ressourcen für jeden einzelnen Verkehrsfluß zu reservieren. Somit können zwischen den NGN nur Ressourcen für aggregierten Verkehr exklusiv vergeben werden.

Ein Protokoll für die Interdomain-Ressourcenreservierung, das die Anforderungen der QoS-Inseln erfüllt, ist das in Abschnitt 4.2 beschriebene NSLP-QoS-Signaling-Protokoll der NSIS-Familie. Aus diesem Grund wird dieses auch im KGID eingesetzt. Allerdings benötigt dieses Protokoll noch eine Schnittstelle zum Interdomain-Routing, auf die im Abschnitt 4.3.3.3 noch näher eingegangen wird.

Mit der Unterteilung der Interdomain-Dienstqualität in QoS-Klassen und der dazugehörigen Ressourcenverwaltung kann prinzipiell die Interdomain-Dienstqualität sichergestellt werden. Dabei entsteht zwischen den NGN eine "Kette des Vertrauens". Mittels der Interdomain-Ressourcenreservierung kann ein NGN überprüfen, ob ein Pfad genügend Ressourcen für die geplante Datenübertragung bereitstellen kann. Aber nach dieser Zusage muß sich ein NGN darauf verlassen, daß die anderen NGN den an sie gesendeten QoS-Verkehr auch ohne Beeinträchtigung der Dienstqualität weiterleiten können.

Um die Interdomain-Dienstqualität in jedem Fall sicherstellen zu können, ist noch eine Anpassung des Interdomain-Routing notwendig und das Verhalten bei Topologieänderungen muß noch untersucht werden.

4.3.2 Einsatz von BGP+ in den QoS-Inseln

Neben der Gewährleistung der Dienstqualität ist innerhalb der QoS-Inseln auch ein stabiles und schnell konvergierendes Interdomain-Routing notwendig, um eine gute Fehlerrobustheit in den QoS-Inseln sicherzustellen. Deswegen wird beim KGID das im letzten Kapitel eingeführte BGP+ als Interdomain-Routing-Protokoll innerhalb der QoS-Inseln eingesetzt. BGP+ muß aber noch angepaßt werden, um die Anforderung der Unterscheidung zwischen den Pfaden, die die Dienstqualität gewährleisten können oder nicht, zu erfüllen.

Dazu ist eine Markierung der Pfade notwendig, die die Dienstqualität sicherstellen können. Das läßt sich am einfachsten, wie bei q-BGP für die Gruppe 1, durch ein optionales Attribut in der BGP+-UPDATE-Nachricht realisieren. Dafür reicht ein einzelnes Bit bei der Übermittlung der BGP+-Routing-Information aus, und die Suche der besten Pfade von BGP+ muß diese zusätzliche Information berücksichtigen. Für diesen Zweck wird in die Bestimmung der besten Pfade aus Abschnitt 2.1.2.2 zwischen der ersten und der zweiten Entscheidungsstufe eine weitere Überprüfung eingefügt. Mit dieser Überprüfung wird entschieden, ob zu dem Präfix QoS-Verkehr geschickt werden kann und somit das Präfix innerhalb der QoS-Insel erreichbar ist. Wenn dies zutrifft, muß das entsprechende Attribut in der BGP+-UPDATE-Nachricht gesetzt sein.

Um sicherzustellen, daß das benachbarte Netz die Dienstqualität zusagen kann, das Attribut setzen kann und das BGP+-Routing unterstützt, sollte dies beim Aufbau jeder BGP+-Verbindung überprüft werden. BGP+ nutzt schon die Möglichkeit eines optionalen Parameters in der BGP+-OPEN-Nachricht, um die Kompatibilität zu BGP zu gewährleisten. Deswegen wird auch für die

Überprüfung der Pfadunterscheidung derselbe Ansatz durch die Definition eines weiteren optionalen Parameters in der BGP+-OPEN-Nachricht gewählt. Die dabei zu beachtenden Punkte wurden schon in Abschnitt 3.6 diskutiert.

Mit diesen Maßnahmen ist sichergestellt, daß der QoS-Verkehr innerhalb der QoS-Inseln bleibt, die Kompatibilität zu BGP am Rand einer QoS-Insel und die Skalierfähigkeit von BGP+ erhalten bleiben. Offen ist aber noch das Routing des BE-Verkehrs innerhalb der QoS-Inseln und dessen Einfluß auf die Ermittlung der besten Pfade von BGP+. Da für den BE-Verkehr die Einschränkungen des QoS-Verkehrs nicht gelten, ist es möglich, daß der BE-Verkehr die QoS-Inseln verlassen und später wieder hineingeleitet werden kann.

Die einfachere Lösung ist die Gleichbehandlung des QoS- und des BE-Verkehrs innerhalb der QoS-Inseln und damit der Ausschluß dieser Option. Dabei ist nur wichtig, daß das Attribut, das die QoS-Unterstützung markiert, in den BGP+-UPDATE-Nachrichten beim Verlassen der QoS-Insel gelöscht wird. Ansonsten sind keine weiteren Maßnahmen notwendig.

Will ein Betreiber die Möglichkeit nutzen, den BE-Verkehr auf anderen Pfaden als den QoS-Verkehr weiterzuleiten, sind weitere Mechanismen notwendig. Die Weiterleitung von Paketen, deren Ziel außerhalb der QoS-Insel liegt, ist davon nicht betroffen. Nur wenn das Ziel des BE-Verkehrs sich innerhalb der QoS-Insel befindet, kann ein BGP+-Router zwischen den Pfaden für QoS-Verkehr und für BE-Verkehr unterscheiden. Dazu wird bei der Paketweiterleitung nicht nur anhand des Ziels, sondern auch anhand der Markierung als BE- oder QoS-Verkehr entschieden, zu welchen Knoten die Pakete weitergeleitet werden. Damit wird dieselbe Voraussetzung getroffen wie in [PSS05, PMR+06], wo das *Type of Service*-Feld des IP-Header auch bei der Paketweiterleitung ausgewertet wird.

Um den BE-Verkehr getrennt vom QoS-Verkehr behandeln zu können, sind zwei getrennte Routing-Tabellen nötig. In der einen Routing-Tabelle werden nur die Pfade für den QoS-Verkehr verwaltet und damit ist diese auf eine QoS-Insel beschränkt. Die zweite Routing-Tabelle beherbergt die besten Pfade für den BE-Verkehr. Für die Pfade der zweiten Routing-Tabelle darf das optionale Attribut mit der Markierung für einen Pfad, der die Dienstqualität gewährleistet, nicht ausgewertet werden. Existieren für ein Präfix der QoS-Insel zwei beste Pfade, dann müssen diese als QoS-Pfad und als BE-Pfad an alle anderen BGP+-Router weitergegeben werden, da nur in diesem Fall sichergestellt ist, daß alle anderen NGN ebenfalls diese Unterscheidung vornehmen können.

Da in diesem Fall unter Umständen zwei beste Pfade zu einem Ziel verwendet werden, muß das optionale Attribut, das die QoS-Pfade markiert, auch Bestandteil des Withdrawn-Teils der BGP+-UPDATE-Nachricht sein. Aufgrund der Integration von EPIC in BGP+ wurde das Format der BGP+-UPDATE-Nachricht schon um diese Möglichkeit erweitert und erfüllt diese Bedingung.

Der Vorteil dieses erhöhten Aufwands ist eine größere Freiheit bei der Auswahl der besten Pfade, wobei der BE-Verkehr Pfade nutzen kann, die keinen QoS-Verkehr transportieren. Allerdings werden dafür zwei Local-RIB, zwei unterschiedliche Algorithmen für die Suche der besten Pfade und zwei FIB in jeder Steckkarte des BGP+-Routers benötigt. Aufgrund dessen steigt der Speicherbedarf. Aber nur für die erreichbaren Präfixe innerhalb einer QoS-Inseln werden mehrere beste Pfade verwaltet. Der Mehraufwand hängt also stark von der Größe der QoS-Insel ab.

Wegen der Auslegung von BGP+ ist es jedem Betreiber eines NGN freigestellt, ob er den Mehraufwand in seinem NGN für die getrennten Pfade für BE- und QoS-Verkehr in Kauf nimmt oder nicht. Somit besteht jeder Zeit die Möglichkeit, diese Unterscheidung wieder aufzuheben. Dies wird vor allem interessant, wenn die QoS-Insel sehr groß wird und damit der resultierende Mehraufwand ein weiteres Wachstum der QoS-Insel behindert. Ein weiterer entscheidender Vorteil dieses Vorgehens ist, daß dabei auch die Skalierbarkeit von BGP+ erhalten bleibt.

4.3.3 Verhalten bei Topologieänderungen

Ein wichtiger Punkt, der bisher noch nicht in Betracht gezogen wurde, ist das Verhalten bei einer Interdomain-Topologieänderung. Kommt es zu einer BGP+-Routing-Änderung, sollten auf dem neu gefundenen Pfad ausreichend Ressourcen für den Transport des QoS-Verkehrs zur Verfügung stehen. Sind zu wenig Ressourcen vorhanden, kann die Dienstqualität nach dieser Routing-Änderung nicht mehr garantiert werden.

Tritt dieser Fall ein, existiert die Möglichkeit einfach einen weiteren neuen Pfad zu suchen und dabei zu hoffen, daß dieser genügend Ressourcen zur Verfügung stellen kann. Ist dies aber wieder nicht erfolgreich, können weitere Pfade auf dieselbe Art und Weise ausprobiert werden. Da ein Netzbetreiber dabei nicht sicher sein kann, ob dies zum Erfolg führt und jedes weitere Umschalten Zeit kostet, in der die Dienstqualität nicht garantiert werden kann, wird eine andere Lösung benötigt [**SE04**]. Um die Lösungsmöglichkeiten einzugrenzen, werden im ersten Schritt die verfügbaren Informationen untersucht.

4.3.3.1 Kenntnisse über verfügbare Ressourcen innerhalb der QoS-Inseln

Im Gegensatz zu den Informationen, die innerhalb eines Netzes zur Verfügung stehen, ist deren Umfang, der zwischen den NGN der QoS-Inseln bekanntgegeben werden kann, deutlich geringer. Der Hauptgrund dafür liegt in den Skalierbarkeitsanforderungen. Aus diesem Grund verteilt BGP+ nur die besten Pfade zu einem Präfix und dies läßt sich auch auf die Informationen über unbenutzte Ressourcen übertragen.

Will ein NGN die aktuell auf allen Pfaden verfügbaren Ressourcen erfahren, muß es für mehr als 225.000 Präfixe die entsprechenden Informationen erfragen. Das stellt für die besten Pfade einmalig kein Problem dar. Aber da sich die freien Ressourcen mit der Zeit ändern, genügt es nicht, diese einmalig zu übertragen. Diese Informationen müssen periodisch erneuert werden. Wird die Periodendauer zu groß gewählt, steigt die Wahrscheinlichkeit, daß eine Information veraltet ist und damit die Wahrscheinlichkeit, daß ein Pfad nicht mehr genügend Ressourcen zur Verfügung stellen kann. Daraus resultiert die Notwendigkeit einer relativ kurzen Periodendauer, was wiederum eine immense Signalisierungslast nach sich zieht, die von den BGP+-Routern nicht bewältigt werden kann.

Aber selbst, wenn man die freien Ressourcen kennen würde, hilft dies im Fall einer Topologieänderung nicht weiter. Wie in Abschnitt 3.3.7 gezeigt wurde, kann das Interdomain-Routing mit

BGP+ nicht vorhergesagt werden. Daraus resultiert, daß erst nachdem alle betroffenen NGN ihren besten Pfad bestimmt haben, bekannt ist, wie sich die Pfade ändern und wieviel Verkehr dabei verschoben wird. Die Konsequenz aus diesem Verhalten ist, daß ein NGN zum Zeitpunkt der Bestimmung seiner besten Pfade nicht weiß, wieviel Verkehr aufgrund dieser Änderung umgelegt wird und somit wie viele freie Ressourcen auf dem neuen besten Pfad benötigt werden.

Somit bringt das Wissen über die aktuell verfügbaren Ressourcen überhaupt keinen Vorteil im Fall einer Topologieänderung. Die Reaktion auf eine Interdomain-Topologieänderung muß ohne die Kenntnisse der verfügbaren Ressourcen auskommen.

4.3.3.2 Konsequenz für die Ressourcenreservierungen

Die Unkenntnis der verfügbaren Ressourcen und die Nichtvorhersagbarkeit der Routing-Entscheidung der einzelnen NGN führt dazu, daß im Fall einer Topologieänderung BGP+ die neuen besten Pfade bestimmt und das QoS-NSLP nur versuchen kann, soviel wie möglich der verfügbaren Ressourcen nach der Topologieänderung zu reservieren. Reichen die Ressourcen nicht aus, kann für die betroffenen Verkehrsflüsse die Dienstqualität nicht mehr sichergestellt werden, und die NGN können keine neue Ressourcenreservierungen für die betroffenen Pfade mehr annehmen.

Um die Dienstqualität trotz dieser Verhaltensweise sicherstellen zu können, müssen die NGN und deren Interdomain-Verbindungen für diesen Fall dimensioniert werden. Dabei müssen zwei Fälle besonders berücksichtigt werden. Zum einen muß das Umlegen von QoS-Verkehr aufgrund einer Topologieänderung von BGP+ mit eingeplant werden und zum zweiten muß ein NGN gewährleisten, daß es den von anderen NGN empfangenen QoS-Verkehr an sein Ziel oder zumindest zum nächsten NGN ohne Einbuße der Dienstqualität weiterleiten kann.

Ungeachtet dieser Auslegung der NGN kann ein Netzbetreiber versuchen, stark ausgelastete Interdomain-Verbindungen zu entlasten, indem er gezielt Verkehr dieser Pfade auf andere Pfade umlegt. Wie dies im heutigen Internet realisiert werden kann, ist ausführlich in [Uhl04, UBQ03] untersucht worden. Dieselben Mechanismen können auch in einem NGN eingesetzt werden. Da die verfügbaren Ressourcen nicht bekannt sind, kann man die in den SLA festgeschriebenen maximal nutzbaren Ressourcen als Anhaltspunkte verwenden. Bezieht man den empfangenen bzw. versendeten Verkehr mit ein und vertraut darauf, daß das Nachbar-NGN die zugesicherte Verkehrsmenge auch weiterleiten kann, erhält man eine gute Abschätzung, um die Verkehrsflüsse an die aktuellen Gegebenheiten anzupassen.

Damit die reservierten Ressourcen nach der Verschiebung von QoS-Verkehr an die neuen Bedingungen angepaßt werden können, ist eine Schnittstelle zwischen BGP+ und dem QoS-NSLP erforderlich. Wie diese Schnittstelle ausgelegt werden muß, wird im folgenden detailliert untersucht.

4.3.3.3 Schnittstelle zum NSIS Transport Layer Protocol

Damit das QoS-NSLP nach einer Interdomain-Topologieänderung die Ressourcenreservierung für den Interdomain-QoS-Verkehr aktualisieren kann, muß es über die Routing-Änderung von BGP+ in Kenntnis gesetzt werden. Wegen des zweistufigen Aufbaus der NSIS-Protokollfamilie muß nicht das QoS-NSLP, sondern das NTLP über die BGP+-Pfadänderung informiert werden (s. Abbildung 4.4).

Sobald das NTLP von der Pfadänderung erfahren hat, sind die weiteren Schritten, um die Interdomain-Ressourcenreservierungen an die neuen Gegebenheiten anzupassen, schon in [SH06] bzw. [HKL+05] definiert. In [SH06] sind auch die fünf folgenden Möglichkeiten spezifiziert, wie das NTLP eine Routing-Änderung erkennen kann.

- Lokaler Auslöser – Hierbei reagiert das NTLP auf die Änderung des Next-Hop-Eintrags in der FIB eines Routers. Allerdings funktioniert dies nur an den Routern, die unmittelbar von der Änderung betroffen sind und auf denen das NTLP läuft.

- Erweiterter Auslöser – Bei dieser Methode greift das NTLP auf die Link-State-Datenbank zu und erkennt somit die Änderungen des Netzes, das in dieser Datenbank abgebildet ist. Dies ist nur realisierbar, wenn innerhalb eines NGN ein Link-State-Routing-Protokoll eingesetzt wird und das NTLP Zugriff auf die Link-State-Datenbank eines Routers erhält.

- GIST-Modus-C-Überwachung – Ändert sich in einem GIST-Paket das *Time-To-Life-Feld* des IP-Headers oder die Schnittstelle, kann ebenfalls auf eine Routing-Änderung geschlossen werden, wie es in [SSL+05] untersucht wurde.

- Überwachung der Datenübertragung – In diesem Fall reagiert das NTLP auf eine Fehlermeldung des NSLP. Das NSLP kann aufgrund einer Veränderung beim Signalisierungsablauf, wie z.B. eine unerwartete Laufzeit oder Abbruch des Datenflusses, eine Fehlermeldung an das NTLP versenden.

- GIST-Erkennung – Die letzte Möglichkeit ist die Fehlererkennung mittels der periodischen Abfrage des Status der benachbarten NTLP-Kommunikationspartner.

Alle fünf Varianten sind nur bedingt geeignet, um eine BGP+-Routing-Änderung zu erkennen und zeitlich korrekt die Aktualisierung der Ressourcenreservierung anzustoßen. Im allgemeinen ändert sich während der Reaktion von BGP+ auf eine Interdomain-Topologieänderung der beste Pfad zu einem Präfix mehrmals. Deswegen führt die sofortige Reaktion auf eine Änderung der FIB zu vorschnellen Reaktionen. Um das zu verhindern, ist die beste Lösung, daß BGP+ den Auslöser für die Reaktion des NTLP definiert und diesen über eine gemeinsame Schnittstelle übermittelt.

Da BGP+ und das NTLP gemeinsam auf einem Router zum Einsatz kommen, reicht eine standardisierte Schnittstelle aus, über die das NTLP ähnlich dem lokalen Auslöser eine Information

von BGP+ bekommt, daß sich eine Routing-Änderung ergeben hat. Diese Anforderung gilt es in die Definition von BGP+ und dem NTLP aufzunehmen.

Neben der Schnittstelle muß in BGP+ noch ein neues Zeitglied definiert werden. Damit kann sichergestellt werden, daß BGP+ erst nach der letzten Änderung des besten Pfads eine Reaktion des NTLP anstößt. Dieses als *Minimum Resource Reservation Interval* (MRRI) bezeichnete Zeitglied wird ähnlich dem MRAI nach jeder Änderung eines besten Pfads gestartet. Wenn das MRRI abgelaufen ist, wird das NTLP über die Pfadänderung informiert. Da die Aktualisierung der Interdomain-Ressourcenreservierung aufgrund der Unkenntnis der verfügbaren Ressourcen nicht zeitkritisch ist, wird als Standardwert für das MRRI 5 Sekunden vorgeschlagen. Trotzdem sollte das MRRI in einem Bereich von einer Sekunde bis zu einigen zehn Sekunden konfigurierbar sein, damit immer sichergestellt werden kann, daß der Wert des MRRI in jedem Fall größer als die Konvergenzzeit von BGP+ ist. Desweiteren sollte das MRRI innerhalb eines NGN immer denselben Wert besitzen.

Mit der korrekten Konfiguration des MRRI und der damit verbundenen Initialisierung durch BGP+ kann die Aktualisierung der Ressourcenreservierungen nach einer Topologieänderung ausgeführt werden, ohne dabei unnötigen Signalisierungsverkehr zu erzeugen. Damit das KGID auch nach einer Topologieänderung die Dienstqualität des Interdomain-QoS-Verkehrs sicherstellen kann, werden im folgenden die Dimensionierungsregeln für die Interdomain-Verbindungen aufgestellt.

4.4 Dimensionierung der Interdomain-Verbindungen

Wie im letzten Abschnitt festgestellt wurde, erfordern die Interdomain-Topologieänderungen eine Überdimensionierung der Interdomain-Verbindungen, so daß diese Verbindungen ausreichend Ressourcen für die Verkehrsverschiebungen des Interdomain-QoS-Verkehrs zur Verfügung stellen und damit die Dienstqualität weiterhin garantiert werden kann.

Wie in Abschnitt 2.2.1 erläutert wurde, ist in heutigen Kernnetzen eine entscheidende Ressource die zur Verfügung stehende Bandbreite. Deswegen wird im folgenden abgeschätzt, wieviel zusätzliche Kapazität auf den Interdomain-Verbindungen bereitgestellt werden muß, um die Dienstqualität auch im Fall einer Interdomain-Topologieänderung zu gewährleisten.

Dazu wurde in [**SG05**] der verschobene Interdomain-Verkehr in verschiedene Kategorien und deren Kombinationen aufgeteilt. Diese sollen anhand des Beispiels in Abbildung 4.6 erklärt werden.

Aufgrund des Ausfalls der Verbindung $< \mathrm{NGN}_1\mathrm{NGN}_2 >$ wird der Interdomain-Verkehrsfluß über den Pfad $< \mathrm{NGN}_1\mathrm{NGN}_2\mathrm{NGN}_4 >$ auf den Ersatzpfad $< \mathrm{NGN}_1\mathrm{NGN}_3\mathrm{NGN}_4 >$ umgelegt. Der einfachste Fall tritt bei NGN_2 auf, wo sich der empfangene Verkehr reduziert. Das Gegenteil passiert bei NGN_3, dieses erhält mehr Verkehr, ohne daß sich etwas an den Verbindungen zu seinen Nachbarn ändert. Bei NGN_1 und NGN_4 bleibt die Gesamtsumme des eintreffenden und ausgehenden Verkehrs pro NGN gleich. Allerdings verläßt bei NGN_1 der betroffene Verkehr

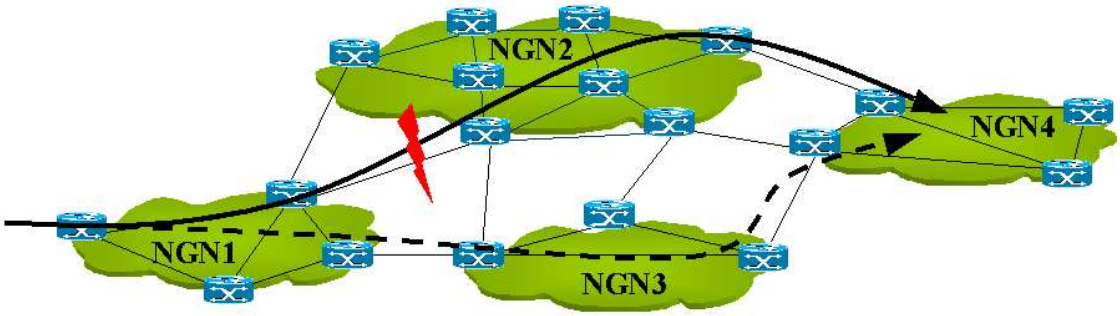


Abbildung 4.6: Interdomain-Verkehrsverschiebungen

über einen anderen Ausgangs-Router das NGN und bei NGN_4 ändert sich der Eingangs-Router für den entsprechenden Verkehr.

Dieses Beispiel zeigt alle möglichen Auswirkungen einer Interdomain-Topologieänderung für ein einzelnes Präfix auf verschiedene NGN. Allerdings sind in der Regel mehrere Präfixe von einer Interdomain-Topologieänderung betroffen. Deswegen setzt sich die Verkehrsverschiebung eines NGN, hervorgerufen durch eine Interdomain-Topologieänderung, aus den Verschiebungen des Verkehrs aller betroffener Präfixe zusammen.

4.4.1 Modellierung

Für die Untersuchung des Einflusses einer Interdomain-Topologieänderung auf den Verkehr, der ein NGN verläßt, passiert oder in ein NGN eintritt, wurde eine Internet-Topologie mit 500 NGN verwendet (s. Abschnitt 2.4.3). Um nicht die Verkehrsverschiebungen aller NGN einzeln auszuwerten, wurden die NGN anhand ihrer Größe in die drei in Abschnitt 2.1.2.4 vorgestellten Kategorien eingeteilt. Alle NGN, die mehr als 79 Nachbarn besitzen, bilden die Gruppe der Tier1-NGN bzw. der Transit-NGN. Die NGN, die nur zwei Interdomain-Verbindungen aufweisen, werden den Tier3-NGN bzw. den Stub-NGN zugeordnet. Als Tier2-NGN werden alle verbleibenden NGN bezeichnet.

Zur Veranschaulichung der Zuordnung der NGN in eine der drei Kategorien ist in Abbildung 4.7 die Verteilung der Anzahl der Nachbar-NGN pro NGN dargestellt.

Die größte Kategorie stellen die Stub-NGN (grüner Balken in Abbildung 4.7) mit mehr als der Hälfte der NGN dar. Eine sehr große Verteilung an Nachbar-NGN weisen die Tier2-NGN (violette Balken in Abbildung 4.7) auf und von Transit-NGN (rote Balken in Abbildung 4.7) gibt es nur sechs. Damit zeigt sich auch das zweite Potenzgesetz aus [FFF99] sehr deutlich (s. Abschnitt 2.4.1). Dieses besagt, daß die Anzahl der Nachbarn eines NGN umgekehrt proportional zu der Häufigkeit der Anzahl an Nachbar-NGN ist. Desweiteren sind die sechs Transit-NGN miteinander vollvermascht und jedes besitzt zu mehr als sechszehn Prozent der restlichen NGN eine direkte Verbindung.

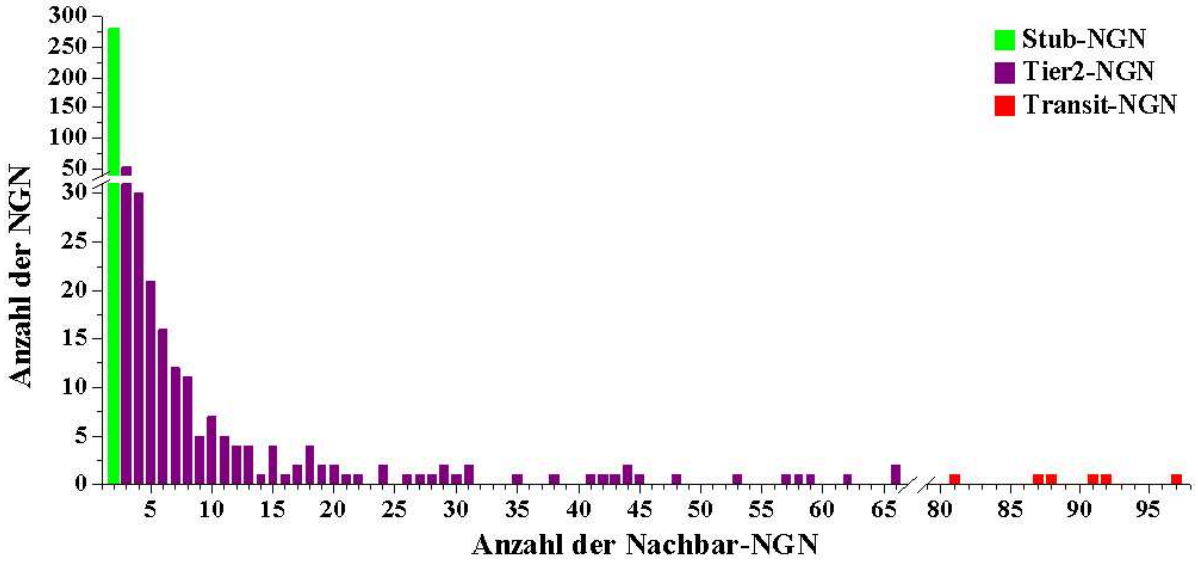


Abbildung 4.7: Verteilung der Anzahl der Nachbar-NGN

Zusätzlich zu der Einteilung der NGN wurden die Interdomain-Verbindungen in Betreiber-, Kunden- und Peer-to-Peer-Verbindungen unterteilt. Als eine Betreiber-Verbindung wird eine Verbindung bezeichnet, die von einem NGN zu dessen Betreiber gerichtet ist. Betreiber eines NGN ist ein anderes Nachbar-NGN, dessen Anzahl an Interdomain-Verbindungen größer ist. Eine Kunden-Verbindung entspricht genau dem Gegenteil, d.h. diese Verbindung ist von einem NGN zu einem seiner Nachbar-NGN gerichtet, das weniger Nachbarn besitzt. Eine Peer-to-Peer-Verbindung entsteht, wenn diese Verbindung zwei Transit-NGN miteinander verbindet oder wenn die beiden verbundenen NGN die gleiche Anzahl an Nachbar-NGN aufweisen.

Diese Einteilung ermöglicht ein BGP-Routing nachzubilden, das die in Abschnitt 2.1.2.3 erklärten Policies aus [Gao01] berücksichtigt. Diese verallgemeinerten Policies schränken die Pfadauswahl zu einem Ziel mit den folgenden Regeln ein:

- Enthält der Pfad eine oder mehrere Kunden-Verbindungen bzw. eine Peer-to-Peer-Verbindung, dürfen nur noch Kunden-Verbindungen zum besten Pfad hinzugefügt werden.
- Eine Peer-to-Peer-Verbindung kann nur ausgewählt werden, wenn der bisherige beste Pfad nur aus Betreiber-Verbindungen besteht oder noch nicht existiert.

Der beste Pfad zu einem Präfix wird mit dem Dijkstra-Algorithmus berechnet, wobei alle Verbindungen dasselbe Verbindungsgewicht besitzen und die für die Berechnung zur Verfügung stehenden Pfade durch die beiden beschriebenen Regeln eingeschränkt werden. Sind zwei Pfade zu einem Ziel gleich lang, dann wird der Pfad bevorzugt, der eine Peer-to-Peer-Verbindung enthält. Reicht diese Eigenschaft für eine Entscheidung nicht aus, wird der Pfad gewählt, der vor der Berechnung die kleinere Identifikationsnummer bekommen hat. Damit erhält man ein

BGP-Routing, das die grundlegenden Policies und die Pfadlänge, gemessen in NGN-Hops, bei der Suche der besten Pfade berücksichtigt.

Aufgrund nicht öffentlich verfügbarer Interdomain-Verkehrsmatrizen wurde eine eigene Verkehrsmatrix erstellt. Bei dieser wird zwischen jedem NGN ein Verkehrsfluß definiert. Die normierte Größe des Verkehrsflusses entspricht der Anzahl der Nachbar-NGN des kleineren NGN. Eine Ausnahme bilden die sechs Transit-NGN. Entsprechend dem Charakter von Transit-NGN leiten diese nur die Verkehrsflüsse der anderen NGN weiter und somit bestehen zwischen den anderen NGN und den Transit-NGN keine direkten Verkehrsflüsse.

Als Ursache der Interdomain-Topologieänderung wurde der Ausfall jeder einzelnen Interdomain-Verbindung untersucht. Dabei wurde eine Interdomain-Verbindung als ausgefallen gekennzeichnet, für alle betroffenen Pfade wurde ein neuer bester Pfad berechnet und die resultierende Verkehrsverschiebung ermittelt. Dieser Ablauf wurde für den Ausfall jeder der 1730 Interdomain-Verbindungen wiederholt.

4.4.2 Ergebnisse

Bei den so ermittelten Ergebnissen wurden für jede Verbindung in beide Richtungen und für jeden Ausfall einer Verbindung die Änderungen des Verkehrsaufkommens abgespeichert. Für die Auswertung wurden nur die positiven Änderungen in Betracht gezogen, da nur hierbei erhöhte Kapazitätsanforderungen entstehen. Negative Änderungen sind mit einer Kapazitätsreduzierung verbunden und werden nicht berücksichtigt.

Aufgrund der großen Anzahl der einzelnen Werte werden die Schlußfolgerungen anhand von Mittelwerten unter Angabe der Konfidenzintervalle für 99,5% diskutiert. Desweiteren wird auch das 99,5%-Quantil in die Betrachtungen miteinbezogen, wobei dann nur die Bezeichnung Quantil verwendet wird. Sollte die Anzahl der relevanten Meßwerte kleiner als zehn sein, werden die Ergebnisse nicht mit betrachtet.

Die Resultate sind für zwei verschiedene Fehlerszenarien ermittelt worden. Das als “direkt” bezeichnete Szenario berücksichtigt alle Ausfälle einer Verbindung des jeweiligen NGN, wohingegen im Szenario “alle“, alle untersuchten Fehlerfälle betrachtet werden. Neben diesen verschiedenen Fehlerbildern wird bei der Erklärung der Ergebnisse die Einteilung der NGN anhand ihrer Größe und die Einteilung der Verbindungen berücksichtigt. Um die Aussagen noch zu verdeutlichen, werden auch die Verkehrsverschiebungen einzelner ausgesuchter NGN vorgestellt.

Da die Eingruppierung der Verbindungen richtungsabhängig ist, werden die verwendeten Bezeichnungen anhand der Abbildung 4.8 aus Sicht eines NGN kurz erklärt.

Bei den Peer-to-Peer-Verbindungen wird nur unterschieden, ob sie von dem betrachteten NGN empfangenen oder gesendeten Verkehr übertragen. Dieselbe Unterscheidung gilt für die restlichen Verbindungen ebenfalls. Allerdings wird bei diesen Verbindungen noch die Kategorie des Nachbar-NGN mitberücksichtigt. Ist eine unidirektionale Verbindung von dem betrachteten NGN zu seinem Betreiber-NGN gerichtet, so wird diese als eine Betreiber-Verbindung mit ausgehendem Verkehr bezeichnet. Überträgt eine Betreiber-Verbindung empfangenen Verkehr, zeigt

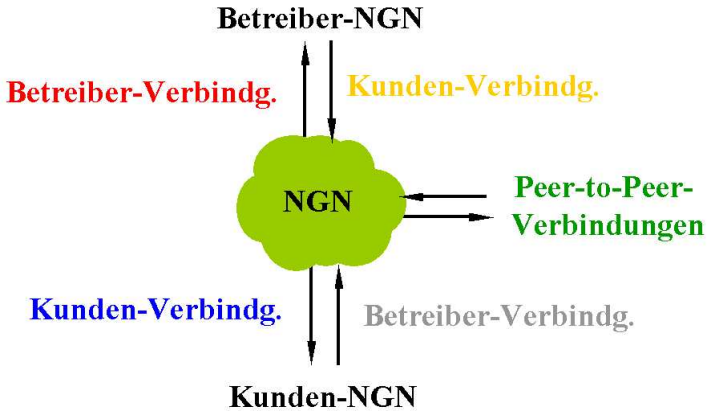


Abbildung 4.8: Bezeichnung der Verbindungen eines NGN

diese Verbindung von einem Kunden-NGN zu dem untersuchten NGN. Eine Kunden-Verbindung mit eingehendem Verkehr stellt eine Verbindung von einem Betreiber-NGN in Richtung des betrachteten NGN dar. Zeigt eine Kunden-Verbindung zu einem Kunden-NGN, kann darüber nur ausgehender Verkehr transportiert werden.

4.4.2.1 Stub-NGN

Ein Stub- oder Tier3-NGN ist ausschließlich eine Quelle oder Senke von Interdomain-Verkehr und hat deswegen im Regelfall nur eine oder zwei Verbindungen zu anderen NGN. In dem untersuchten Modell hat jedes der Stub-NGN genau zwei Interdomain-Verbindungen, da sonst das NGN bei einem Ausfall der einzigen Verbindung vom restlichen Netz getrennt würde. Wegen dieser zwei Verbindungen kommt es zu sehr starken Verkehrsverschiebungen, wenn eine der beiden Verbindungen ausfällt.

Im Extremfall ändert sich der Verkehr um den Faktor 492 im Vergleich zum fehlerfreien Szenario. Dabei fällt die Verbindung aus, die im fehlerfreien Fall fast den kompletten Verkehr transportiert. Resultierend aus der Definition der Verkehrsmatrix hat jedes NGN genau 493 Verkehrsbeziehungen zu den anderen NGN. Davon ausgenommen sind die sechs Transit-NGN und das betrachtete NGN. Desweiteren führt die Definition der Verkehrsmatrix dazu, daß ein Stub-NGN $493 * 2 = 986$ Kapazitätseinheiten (KE) versendet bzw. empfängt.

Betrachtet man den Extremfall, werden $492 * 2\mathrm{KE} = 984\mathrm{KE}$ über die eine Verbindung und über die zweite 2KE übertragen. Dabei transportiert die zweite Verbindung genau den Verkehr, der exakt der Verkehrsbeziehung zu dem Nachbar-NGN entspricht. Fällt nun die Verbindung, die den Großteil des Verkehrs trägt, aus, muß die zweite Verbindung den kompletten Verkehr transportieren. Das entspricht einer Erhöhung von 2KE auf 986KE und damit erhöht sich der Kapazitätsbedarf um den Faktor 492.

Da dieser Extremfall nur neunmal auftritt, liegt die mittlere Verkehrsverschiebung deutlich darunter, wie in Tabelle 4.1 zu sehen ist.

Stub-NGN	Art	$\varnothing_{in}$	$\varnothing_{out}$
Kunden-Verbindg.	direkt	$2.041,7 \pm 523,0\%$	$k.A.$
	alle	$45,0 \pm 11,0\%$	$k.A.$
Betreiber-Verbindg.	direkt	$k.A.$	$672,2 \pm 270,9\%$
	alle	$k.A.$	$9,8 \pm 3,2\%$

Tabelle 4.1: Verkehrsverschiebungen der Stub-NGN

Zieht man nur den Ausfall jeweils einer Verbindung eines Stub-NGN in Betracht, erhöht sich der eingehende Verkehr im Mittel auf 2041%. Dagegen liegt die mittlere Änderung beim ausgehenden Verkehr nur bei 672%. Dies verdeutlicht die asymmetrische Verteilung des Verkehrs auf die beiden Verbindungen eines Stub-NGN, was auch anhand der Schwankungen für alle Fehler sichtbar wird.

Jedes NGN sucht unabhängig von den anderen NGN anhand der verallgemeinerten Policies und der Länge des Pfads seinen besten Pfad. Es läßt sich nicht vermeiden, daß der Pfad zwischen zwei NGN sich je nach Richtung unterscheidet. Im Fall der Stub-NGN führt dies zu einer deutlichen Konzentration des ausgehenden Verkehrs auf eine Verbindung, wohingegen auf der Rückrichtung der empfangene Verkehr besser auf die beiden Verbindungen der Stub-NGN verteilt wird.

Die Auswirkungen dieser Asymmetrie und der sehr starken Verkehrsschwankungen werden sich auch in den folgenden Ergebnissen für die Tier2- und Transit-NGN wiederfinden.

4.4.2.2 Kleine Tier2-NGN

Die Tier2-NGN bieten im Gegensatz zu den Transit- und Stub-NGN alle drei verschiedenen Arten von Verbindungstypen und sind selbst Quelle bzw. Senke von Verkehr. Außerdem unterscheiden sie sich auch sehr stark in der Anzahl ihrer Nachbar-NGN, die von drei bis neunundsiebzig reicht. Deswegen werden die Tier2-NGN nochmals unterteilt. Die als kleine Tier2-NGN bezeichneten NGN besitzen zwischen drei und zehn Nachbarn. Dagegen sind alle NGN mit mehr als zehn und weniger als achtzig Nachbar-NGN große Tier2-NGN, deren Ergebnisse im nächsten Abschnitt behandelt werden.

Die mittleren Verkehrsschwankungen der kleinen Tier2-NGN sind in der Tabelle 4.2 aufgeführt.

Tier2-NGN, klein	Art	$\varnothing_{in}$ mit Stub	$\varnothing_{out}$ mit Stub
Peer-to-Peer-Verbindg.	alle	$7,4 \pm 2,2\%$	$7,4 \pm 2,2\%$
Kunden-Verbindg.	direkt	$115,3 \pm 28,5\%$	$k.A.$
	alle	$6,4 \pm 1,0\%$	$130,9 \pm 39,1\%$
Betreiber-Verbindg.	direkt	$k.A.$	$61,5 \pm 31,6\%$
	alle	$35,4 \pm 14,2\%$	$4,0 \pm 1,0\%$

Tabelle 4.2: Verkehrsverschiebungen der Tier2-NGN mit bis zu 10 Nachbarn

Bei diesen Werten fallen sofort die starken Verkehrsverschiebungen der Kunden-Verbindungen auf. Das Quantil des ausgehenden Verkehrs beträgt 170% und verdeutlicht die starken Schwankungen von den kleinen Tier2- zu den Stub-NGN. In der Gegenrichtung wird auf den Betreiber-Verbindungen mit eingehendem Verkehr die in Abschnitt 4.4.2.1 erwähnte Asymmetrie deutlich, da hier das Quantil bei nur 49,6% liegt.

Die Änderungen zu den Betreiber-NGN sind mit im Mittel 6,4% und 4,0% im Vergleich sehr gering, da hier der von den kleinen Tier2-NGN erzeugte bzw. terminierte Verkehr die Schwankungen der Stub-NGN überlagert. Allerdings wachsen diese stark an, wenn nur die direkten Ausfälle berücksichtigt werden. Dabei weisen die Quantile mit 143,8% bzw. 93,1% dieselbe Asymmetrie wie die Stub-NGN bzw. die Verbindungen zwischen den kleinen Tier2-NGN und den Stub-NGN auf. Allerdings werden diese Änderungen von den Verkehrsflüssen der kleinen Tier2-NGN verursacht.

Bei den Verkehrsverschiebungen der Peer-to-Peer-Verbindungen sind diese für den ein- und ausgehenden Verkehr gleich. Aufgrund der Definition in Abschnitt 4.4.1 sind zwei NGN mittels einer Peer-to-Peer-Verbindung miteinander verbunden, wenn die Anzahl der Nachbarn beider NGN gleich ist. Das hat zur Folge, daß relativ wenige Peer-to-Peer-Verbindungen existieren und das damit nur NGN, die der gleichen Kategorie angehören, miteinander verbunden sind. Ändert sich der ausgehende Verkehr eines NGN auf einer Peer-to-Peer-Verbindung, so erfährt das andere NGN dieselbe Änderung bei dessen eingehendem Verkehr. Somit wird dieselbe Änderung beim ein- und ausgehenden Verkehr der NGN mitgezählt und es reicht aus, einen der beiden Verkehre zu betrachten. Da bei jedem NGN andere direkte Fehler betrachtet werden, gilt dieser Zusammenhang für das Fehlerszenario nicht, wo nur die Ausfälle einer direkten Verbindung des NGN berücksichtigt werden.

Die Ergebnisse in Tabelle 4.3 stellen die Verkehrsverschiebungen dar, wenn die Schwankungen der Stub-NGN nicht betrachtet werden. Dabei werden die Änderungen des Verkehrs, verursacht durch den Ausfall einer Verbindung eines Stub-NGN, bei den Berechnungen nicht mitgezählt.

Tier2-NGN, klein	Art	$\varnothing_{in}$ ohne Stub	$\varnothing_{out}$ ohne Stub
Peer-to-Peer-Verbindg.	alle	$4,7 \pm 1,0\%$	$4,7 \pm 1,0\%$
Kunden-Verbindg.	direkt	$122,4 \pm 30,2\%$	$k.A.$
	alle	$10,6 \pm 1,8\%$	$25,1 \pm 9,0\%$
Betreiber-Verbindg.	direkt	$k.A.$	$63,4 \pm 32,5\%$
	alle	$16,7 \pm 9,6\%$	$6,3 \pm 1,8\%$

Tabelle 4.3: Verkehrsverschiebungen der kleinen Tier2-NGN ohne Berücksichtigung der Stub-NGN

Ohne den Einfluß der Verkehrsschwankungen der Stub-NGN reduzieren sich die Verschiebungen auf den Verbindungen zwischen den kleinen Tier2-NGN zu dessen Kunden im Mittel um 105,8% beim ausgehenden und um 18,7% beim eingehenden Verkehr. Diese drastische Verringerung zeigt sehr deutlich den Einfluß der Stub-NGN, und am Unterschied zwischen den beiden Werten wird wieder die Asymmetrie dieses Einflusses deutlich.

Bei den Verschiebungen auf den Verbindungen zu den Betreiber-NGN, also auf den ausgehenden Betreiber- und eingehenden Kunden-Verbindungen, zeigt sich der geringe Einfluß der Stub-NGN, da sich hier die Werte sogar um 2,3% bzw. 4,2% erhöhen. Der Grund für diese Verkehrsschwankungen liegt an den Verkehrsflüssen, die bei den kleinen Tier2-NGN beginnen bzw. enden. Überdeutlich treten diese Verschiebungen in den Vordergrund, wenn nur die direkten Ausfälle betrachtet werden. Bei diesen Änderungen läßt sich auch wieder eine Asymmetrie feststellen, die auf demselben Verhalten begründet ist wie bei den Stub-NGN (s. Abschnitt 4.4.2.1).

Dasselbe Verhalten wird auch bei den Verkehrsverschiebungen von NGN82 deutlich, welche in Abbildung 4.9 unter der Berücksichtigung der Stub-NGN dargestellt sind.

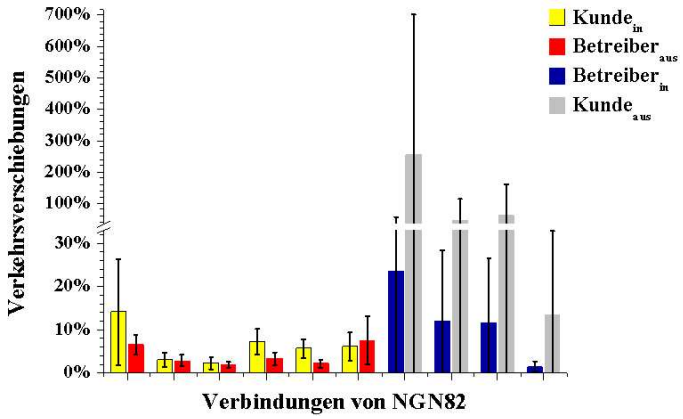


Abbildung 4.9: Verkehrsverschiebungen bei NGN82 mit Stub-NGN

Sechs der zehn Interdomain-Verbindungen von NGN82 verknüpfen dieses mit seinen Betreiber-NGN. Deren Schwankungen (gelbe und rote Balken) liegen in der Regel im einstelligen Prozentbereich, wobei diese Mittelwerte bei Nichtbetrachtung der Stub-NGN um bis zu 2,1% größer sind. Nur auf der Verbindung von NGN34 zu NGN82 kommt es zu wesentlich größeren Verschiebungen, die für alle Fehlerszenarien bei $14,1 \pm 12,2\%$ und ohne die Ausfälle der Verbindungen der Stub-NGN bei $19,3 \pm 20,0\%$ liegen. Somit wird im Fehlerfall ein relativ großer Teil des von NGN82 terminierten Verkehrs auf diese Verbindung umgelegt.

Wesentlich größere Verkehrsverschiebungen treten bei den Verbindungen zu den Kunden-NGN auf. Wobei in diesem Fall zwischen zwei Verhaltensweisen unterschieden werden muß. Die beiden Verbindungen, die in Abbildung 4.10 die kleineren Schwankungen aufweisen, führen zu Stub-NGN, denen NGN82 einen Ersatzpfad anbietet. Dementsprechend wird viel Verkehr nach einem Fehler darauf umgelegt, wobei sich dieser drastisch verringert, wenn die Änderungen des Verkehrs der Stub-NGN nicht berücksichtigt werden.

Die beiden anderen Verbindungen mit den deutlich größeren Schwankungen verbinden NGN82 mit zwei weiteren kleinen Tier2-NGN. Deren eingehender Verkehr wird im Fehlerfall über NGN82 geleitet und daraus resultieren die drastischen Schwankungen. Allerdings ist die Asymmetrie des Verkehrs bei dem NGN mit den größten Verkehrsverschiebungen stärker als bei dem zweiten.

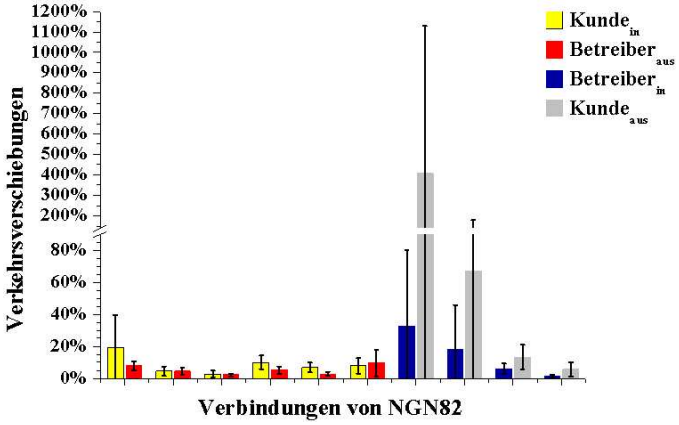
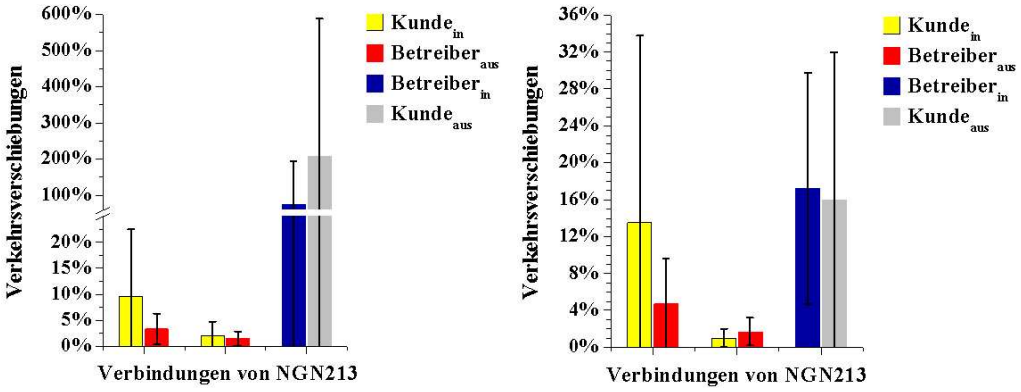


Abbildung 4.10: Verkehrsverschiebungen bei NGN82 ohne Stub-NGN

Ein Teil dieses Verhaltens läßt sich auch bei NGN213 feststellen, wobei dieses NGN nur drei Verbindungen besitzt und deren Verkehrsschwankungen in Abbildung 4.11 zu sehen sind.



(a) (b)

Abbildung 4.11: Verkehrsverschiebungen bei NGN213 mit (a) bzw. ohne (b) Stub-NGN

Bei den Verbindungen mit den beiden Betreiber-NGN existiert ebenfalls eine Verbindung, bei der der empfangene Verkehr stärkere Schwankungen aufweist. Der Grund dafür ist die größere Beanspruchung im Fehlerfall. Der dabei umgelegte Verkehr ist aber hauptsächlich von NGN213 terminierter Verkehr, da die Verschiebungen stärker ausfallen, wenn der Einfluß der Stub-NGN nicht berücksichtigt wird, wie in Abbildung 4.11 (b) zu sehen ist. Im Gegensatz dazu kann bei der Verbindung zu einem Stub-NGN eine drastische Reduzierung der Verkehrsänderungen beobachtet werden, wenn dessen Verkehrsverschiebungen nicht in Betracht gezogen werden.

4.4.2.3 Große Tier2-NGN

Nach der Untersuchung der Verkehrsverschiebungen der kleinen Tier2-NGN werden im folgenden die Schwankungen des Verkehrs der großen Tier2-NGN, die zwischen 11 und 79 Nachbarn besitzen, diskutiert. Über alle großen Tier2-NGN gemittelt sind die Durchschnittswerte der Verkehrsänderungen der einzelnen Verbindungstypen in Tabelle 4.4 dargestellt.

Tier2-NGN, groß	Art	$\varnothing_{in}$ mit Stub	$\varnothing_{out}$ mit Stub
Peer-to-Peer-Verbindg.	direkt	$8,2 \pm 2,0\%$	$6,3 \pm 1,7\%$
	alle	$3,6 \pm 0,4\%$	$3,6 \pm 0,4\%$
Kunden-Verbindg.	direkt	$13,9 \pm 1,0\%$	$8,2 \pm 0,4\%$
	alle	$4,1 \pm 0,1\%$	$8,7 \pm 0,9\%$
Betreiber-Verbindg.	direkt	$12,2 \pm 0,6\%$	$8,8 \pm 0,4\%$
	alle	$4,1 \pm 0,2\%$	$3,8 \pm 0,1\%$

Tabelle 4.4: Verkehrsverschiebungen der Tier2 NGN mit mehr als 10 Nachbarn

Gemittelt über alle Fehlerfälle liegen die Verkehrsverschiebungen für alle drei Verbindungsarten zwischen 3% und 4%. Die einzige Ausnahme stellt der versendete Verkehr über die Kunden-Verbindungen dar. Dessen Mittelwert beträgt 8,7% und ist damit mehr als doppelt so groß. Den Grund dafür stellt die asymmetrische Verkehrsverteilung auf den Verbindungen der Stub-NGN dar. Allerdings ist dieser Einfluß bei weitem nicht mehr so stark ausgeprägt wie bei den kleinen Tier2-NGN, da der Verkehr der Stub-NGN von dem von den großen Tier2-NGN erzeugten bzw. terminierten Verkehr sowie dem Transit-Verkehr überlagert wird.

Betrachtet man nur die Ausfälle einer Verbindung, die direkt mit dem entsprechenden großen Tier2-NGN verbunden ist, steigen die Schwankungen deutlich. Eine Ausnahme bildet wiederum der über die Kunden-Verbindungen versendete Verkehr. Hier verringern sich die Verkehrsverschiebungen, da die asymmetrischen Schwankungen der Stub-NGN sowie der kleinen Tier2-NGN nicht mehr so stark ins Gewicht fallen, weil diese nur noch teilweise direkt mit dem betrachteten großen Tier2-NGN verbunden sind.

Für die Peer-to-Peer-Verbindungen gelten dieselben Aussagen wie für die kleinen Tier2-NGN.

Welchen Einfluß die Verkehrsverschiebungen der Stub-NGN besitzen, zeigt sich in den Ergebnissen in Tabelle 4.5, wo die Änderungen des Verkehrs der Stub-NGN nicht berücksichtigt wurden.

Dabei erhöhen sich die Verkehrsschwankungen im Mittel bei Betrachtungen aller Fehler bei den Peer-to-Peer-Verbindungen um $0,3\%$ und bei den anderen zwischen 0,6% und 1,0%. Eine Ausnahme bildet wieder der zu den Kunden-NGN gesendete Verkehr, dessen Änderungen um 0,7% abnehmen. Auch daran wird deutlich, daß die Verkehrsverschiebungen verursacht durch die Stub-NGN in den Hintergrund treten, und wie die allgemeine Erhöhung zeigt, sogar im Mittel zu niedrigeren Schwankungen führen. Nur beim über die Kunden-Verbindungen versendeten Verkehr ist der negative Einfluß der Verschiebungen von Verkehr der Stub-NGN im Mittel noch sichtbar.

Tier2-NGN, groß	Art	$\varnothing_{in}$ ohne Stub	$\varnothing_{out}$ ohne Stub
Peer-to-Peer-Verbindg.	alle	$3,9 \pm 0,5\%$	$3,9 \pm 0,5\%$
Kunden-Verbindg.	direkt	$14,3 \pm 1,1\%$	$7,9 \pm 0,4\%$
	alle	$5,1 \pm 0,2\%$	$8,0 \pm 0,6\%$
Betreiber-Verbindg.	direkt	$12,0 \pm 0,6\%$	$9,0 \pm 0,4\%$
	alle	$4,7 \pm 0,2\%$	$4,5 \pm 0,1\%$

Tabelle 4.5: Verkehrsverschiebungen der großen Tier2-NGN ohne Berücksichtigung der Stub-NGN

Da die Anzahl der Nachbar-NGN bei den Tier2-NGN so stark variiert, werden im folgenden die Verkehrsverschiebungen anhand von zwei Beispiel-NGN einzeln untersucht. Das NGN10 besitzt 44 Nachbar-NGN und die Schwankungen auf allen Verbindungen von NGN10 sind in Abbildung 4.12 veranschaulicht.

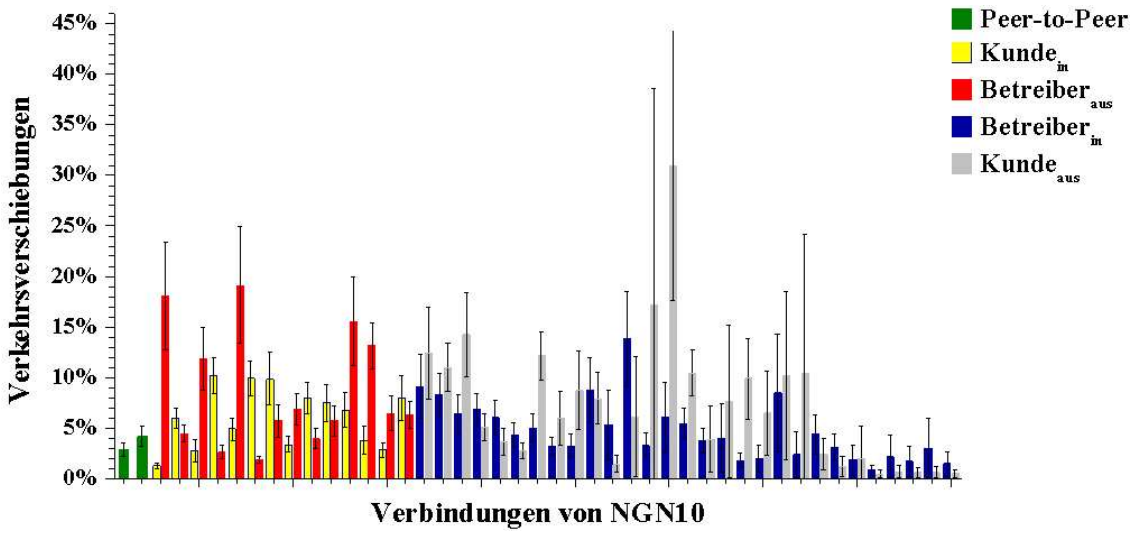


Abbildung 4.12: Verkehrsverschiebungen bei NGN10 mit Stub-NGN

Die in rot dargestellten Schwankungen betreffen den zu den Betreiber-NGN versendeten Verkehr, deren Quantil bis auf 24,9% steigt. Diese Verkehrsänderungen setzen sich zusammen aus den Schwankungen des von den Kunden-NGN empfangenen Verkehrs (blauen Balken) sowie den Verschiebungen der von NGN10 erzeugten Verkehrsflüsse, die über die Betreiber-NGN weitergeleitet werden. Betrachtet man die mittleren Verschiebungen von $4,0 \pm 0,4\%$ für den eingehenden und von $7,0 \pm 0,6\%$ für den ausgehenden Verkehr, kann man feststellen, daß die Schwankungen des von NGN10 erzeugten Verkehrs den größeren Einfluß auf die Verkehrsverschiebungen zu den Betreiber-NGN besitzen.

In der Gegenrichtung, d.h. bei dem von den Betreiber-NGN empfangenen und zu den Kunden-NGN weitergeleiteten Verkehr, spielen andere Einflüsse die entscheidende Rolle. Hier unterscheiden sich die mittleren Schwankungen mit $4,7 \pm 0,4\%$ für den eingehenden und mit

$5,3 \pm 1,1\%$ für den ausgehenden Verkehr deutlich weniger. Allerdings können die Änderungen bei dem zu den Kunden-NGN gesendeten Verkehr bei einzelnen Verbindungen wesentlich stärker ausfallen. Dabei handelt es sich um den von den kleinen Tier2-NGN erzeugten und im Fehlerfall umgelegten Verkehr. Dessen Schwankungen sind auch bei den Änderungen des zu den Betreiber-NGN versendeten Verkehrs bei NGN83 und NGN213 in Abschnitt 4.4.2.2 festgestellt worden.

Wie schon bei den über alle großen Tier2-NGN gemittelten Ergebnisse zeigt sich auch in Abbildung 4.13, daß die Verkehrsschwankungen der Stub-NGN nur noch bedingt einen Einfluß besitzen.

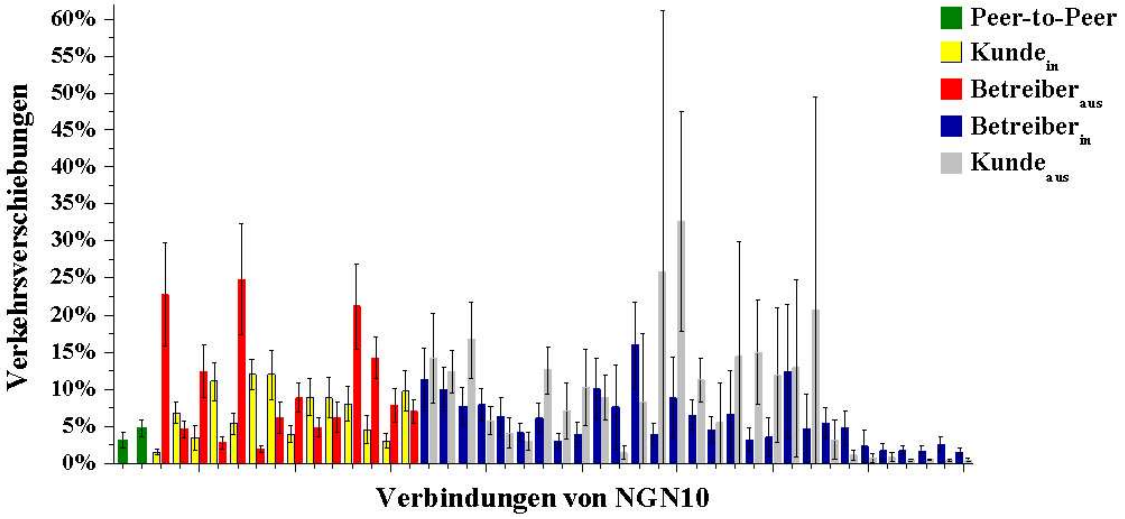


Abbildung 4.13: Verkehrsverschiebungen bei NGN10 ohne Stub-NGN

Bei den im Vergleich kleinen und mittleren Verkehrsverschiebungen läßt sich eine kleine Reduktion erkennen, wenn die Effekte der Stub-NGN nicht berücksichtigt werden. Allerdings verstärken sich durch die Nichtbeachtung der Stub-NGN die schon großen Schwankungen noch. Womit nochmals deutlich wird, daß die Änderungen der Verkehrsflüsse der Stub-NGN sich den Verkehrsverschiebungen des von den Tier2-NGN erzeugten bzw. terminierten Verkehrs unterordnen.

Bei der einzigen Peer-to-Peer-Verbindung von NGN10 bewegen sich die Verkehrsverschiebungen in demselben Bereich wie der Mittelwert für alle großen Tier2-NGN, und auch die Nichtberücksichtigung der Schwankungen der Stub-NGN führt nicht zu erwähnenswerten Änderungen.

Als weiteres Beispiel soll noch ein mittelgroßes Tier2-NGN, das NGN41, betrachtet werden, dessen Verkehrsverschiebungen verteilt auf die 18 Verbindungen in Abbildung 4.14 zu sehen sind, wo die Schwankungen der Stub-NGN enthalten sind.

NGN41 zeigt zum Teil ein etwas anderes Verhalten als NGN10, welches deutlich mehr Nachbarn besitzt. Keinen Unterschied gibt es bei der Peer-to-Peer-Verbindung, deren Quantile bei 5% bzw. darunter liegen.

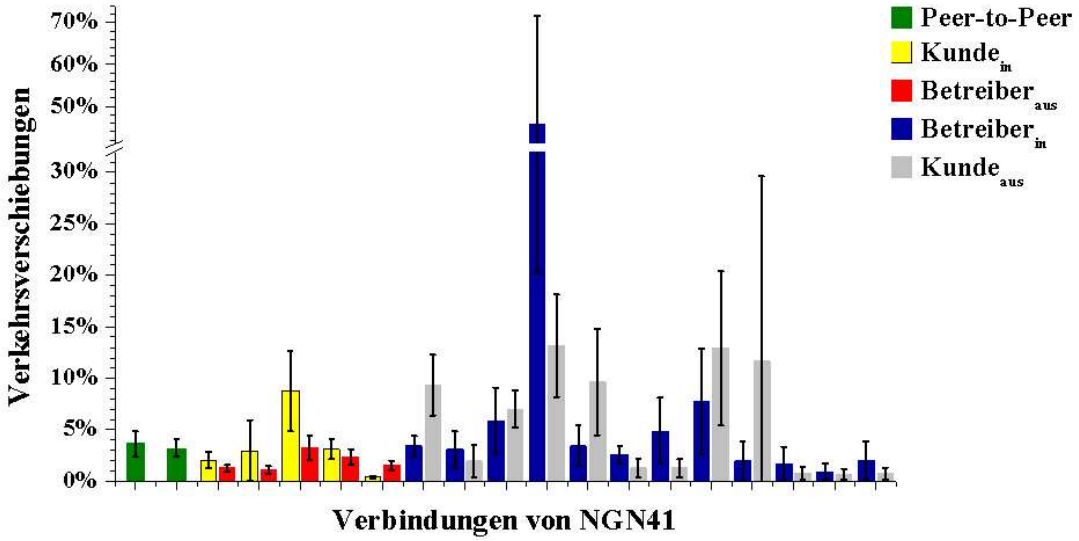


Abbildung 4.14: Verkehrsverschiebungen bei NGN41 mit Stub-NGN

Bei den Verbindungen zu den Betreiber-NGN kehren sich die Ergebnisse von NGN10 um, und die Verkehrsverschiebungen sind deutlich geringer. Die Schwankungen des zu den Betreiber-NGN gesendeten Verkehrs ändern sich im Gegensatz zu NGN10 schwächer als die der von den Betreiber-NGN empfangenen Verkehrs.

Auf den Verbindungen zu den Kunden-NGN fällt eine Verkehrsverschiebung, dessen Quantil 71,6% beträgt, aus dem Rahmen. Hierbei handelt es sich um die Umlegung des von NGN145 erzeugten Verkehrs, der die auch bei den Stub-NGN auftretende Asymmetrie aufweist. Da NGN145 als ein kleines Tier2-NGN klassifiziert ist, tritt diese starke Schwankung auch in Abbildung 4.15 auf.

Die geringere Größe von NGN41 im Vergleich zu NGN10 wird auch deutlich, indem der Einfluß der Verkehrsverschiebungen der Stub-NGN zunimmt. Die in Abschnitt 4.4.2.1 beschriebene Asymmetrie zeigt sich bei der Abnahme der mittleren Schwankungen des zu NGN288 gesendeten Verkehrs von $11,7 \pm 18,0\%$ auf $3,0 \pm 1,4\%$.

4.4.2.4 Transit-NGN

Ein Transit- oder Tier1-NGN stellt das genaue Gegenteil zu einem Stub-NGN dar. Nicht nur, daß die Transit-NGN die größten NGN des gesamten Internets sind, sondern sie leiten nur den von anderen NGN empfangenen Verkehr an ihre Nachbarn weiter. Wegen des ausschließlichen Transports von Transit-Verkehr ist ein Transit-NGN weder Quelle noch Senke von Verkehr. Wie schon in Abschnitt 2.1.2.3 erläutert wurde, besitzt ein Transit-NGN nur Peer-to-Peer-Verbindungen zu den anderen Transit-NGN und ist ein Betreiber für die restlichen Nachbar-NGN. Dies zeigt sich auch in Tabelle 4.6 mit den Verkehrsverschiebungen gemittelt über alle Transit-NGN.

Grundsätzlich schwankt der Verkehr der Peer-to-Peer- und der Betreiber-Verbindungen mit rund

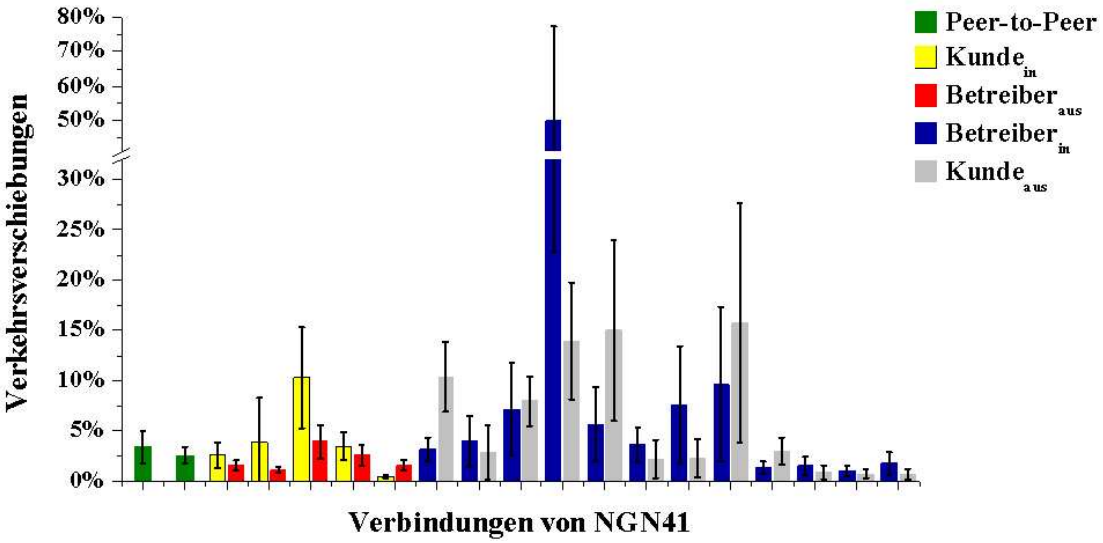


Abbildung 4.15: Verkehrsverschiebungen bei NGN41 ohne Stub-NGN

Transit-NGN	Art	$\varnothing_{in}$ mit Stub	$\varnothing_{out}$ mit Stub
Peer-to-Peer-Verbindg.	direkt	$3,8 \pm 0,2\%$	$4,2 \pm 0,4\%$
	alle	$2,0 \pm 0,1\%$	$2,0 \pm 0,1\%$
Kunden-Verbindg.	direkt	$k.A.$	$7,4 \pm 0,4\%$
	alle	$k.A.$	$3,3 \pm 0,2\%$
Betreiber-Verbindg.	direkt	$11,2 \pm 0,5\%$	$k.A.$
	alle	$2,2 \pm 0,1\%$	$k.A.$

Tabelle 4.6: Verkehrsverschiebungen der Transit-NGN

2,0% auf einem sehr niedrigen Niveau. Allerdings weicht auch hier der über die Kunden-Verbindungen gesendete Verkehr mit einem Mittelwert von 3,3% von den restlichen mittleren Schwankungen ab. Damit ist auch hier die Asymmetrie der Verkehrsverschiebungen der Stub- und kleinen Tier2-NGN feststellbar.

Bei den direkten Ausfällen ist eine deutliche Vergrößerung der Schwankungen festzustellen, wobei hier die Änderungen des empfangenen Verkehrs mit einem Quantil von 11,7% wesentlich größer sind als beim gesendeten Verkehr, wo das Quantil bei 7,8% liegt. Das spiegelt genau das Verhalten der Schwankungen der kleinen und großen Tier2-NGN auf den Verbindungen zu deren Betreiber-NGN wider, die zum Großteil Transit-NGN sind.

Wie bei den anderen NGN auch unterscheiden sich die Verkehrsverschiebungen der Peer-to-Peer-Verbindungen nicht zwischen dem ein- und ausgehenden Verkehr, wenn alle Fehler betrachtet werden. Da bei einem Transit-NGN der Verkehr dieses NGN nur passiert, gleicht der ausgehende Verkehr auf den Peer-to-Peer-Verbindungen den Unterschied zwischen den Schwankungen der Betreiber- und Kunden-Verbindungen bei einem direkten Ausfall aus, was zu einer Erhöhung der Schwankungen bei den ausgehenden Peer-to-Peer-Verbindungen führt.

Bezieht man die Verkehrsverschiebungen verursacht durch Ausfälle bei den Stub-NGN nicht in die Betrachtungen mit ein, fallen die Änderungen der Mittelwerte für die Transit-NGN klein aus, wie aus der Tabelle 4.7 ersichtlich wird.

Transit-NGN	Art	$\varnothing_{in}$ ohne Stub	$\varnothing_{out}$ ohne Stub
Peer-to-Peer-Verbindg.	direkt	$3,9 \pm 0,3\%$	$4,2 \pm 0,4\%$
	alle	$1,8 \pm 0,1\%$	$1,8 \pm 0,1\%$
Kunden-Verbindg.	direkt	$k.A.$	$7,0 \pm 0,4\%$
	alle	$k.A.$	$3,9 \pm 0,2\%$
Betreiber-Verbindg.	direkt	$11,0 \pm 0,5\%$	$k.A.$
	alle	$2,9 \pm 0,1\%$	$k.A.$

Tabelle 4.7: Verkehrsverschiebungen der Transit-NGN ohne Berücksichtigung der Stub-NGN

Bei den Verbindungen zu den Kunden-NGN erhöhen sich die Mittelwerte für alle Fehlerfälle um 0,6% bzw. 0,7%, wobei sich die Änderungen auf den Peer-to-Peer-Verbindungen marginal verringern. Genau wie bei den großen Tier2-NGN gilt für die Transit-NGN, daß der von den Tier2-NGN erzeugte bzw. terminierte Verkehr die Verkehrsverschiebungen wesentlich stärker beeinflußt als die Schwankungen der Stub-NGN.

Die bisher diskutierten Werte beziehen sich auf die Mittelwerte aller sechs Transit-NGN. Um die Verkehrsverschiebungen auch für ein einzelnes NGN zu beurteilen, werden im folgenden die Schwankungen des Verkehrs des größten NGN, dem NGN3 mit 97 Nachbar-NGN näher untersucht.

In Abbildung 4.16 sind die Verkehrsverschiebungen der einzelnen Verbindungen von NGN3 aufgrund aller betrachteten Fehlerfälle dargestellt.

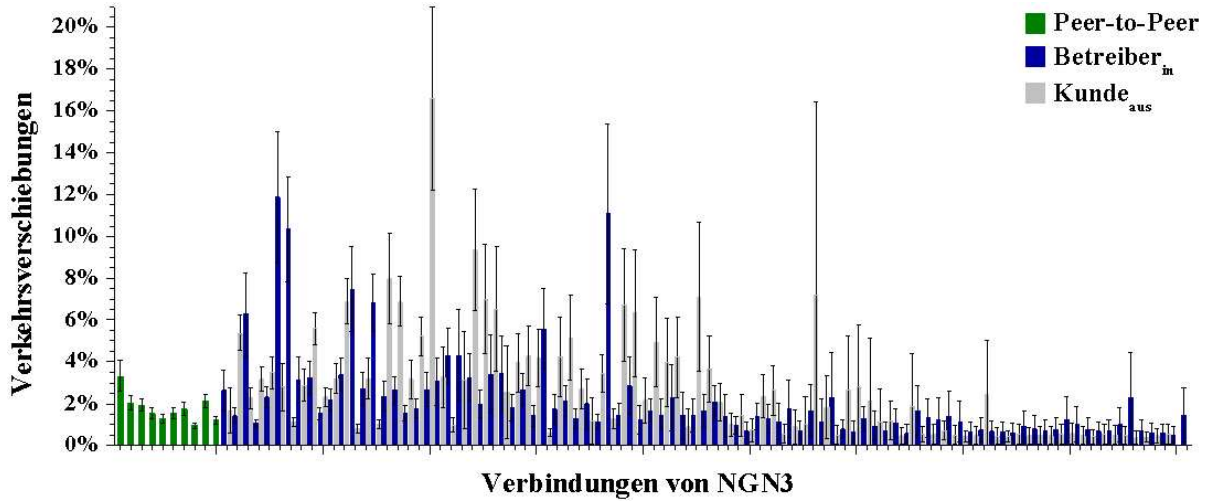


Abbildung 4.16: Verkehrsverschiebungen von NGN3

Hierbei wird deutlich, daß die Änderungen einer einzelnen Verbindung zu einem Kunden-NGN deutlich größer ausfallen können, als der Mittelwert aller Transit-NGN in Tabelle 4.6 vermuten läßt. Die maximale Verkehrsverschiebung einer einzelnen Verbindung von NGN3 mit einem Kunden-NGN beträgt bei 99,5% aller Fehlerfälle 15,4% beim empfangenen und 21,0% beim gesendeten Verkehr. Bildet man den Durchschnitt für alle Verbindungen, liegen die Mittelwerte nur noch bei $1,9 \pm 0,1\%$ für den ein- bzw. bei $2,4 \pm 0,2\%$ für den ausgehenden Verkehr. Damit werden auch die Erkenntnisse aus den Werten von Tabelle 4.7 bestätigt, daß der ausgehende Verkehr stärkeren Schwankungen unterworfen ist als der eingehende.

Schaut man sich einmal an, wo die größten Schwankungen auftreten, stellt man fest, daß alle Schwankungen, deren Quantil größer als 5% ist, bei den Verbindungen mit großen Tier2-NGN auftreten. Beim empfangenen Verkehr betrifft dies 14 Verbindungen, und beim gesendeten Verkehr ist das Quantil der Verschiebungen bei 28 Verbindungen größer als 5%.

Ein ebenfalls asymmetrisches Verhalten kann bei den Peer-to-Peer-Verbindungen festgestellt werden. Deren Mittelwerte liegen bei $1,4 \pm 0,1\%$ für den ein- und bei $2,1 \pm 0,2\%$ für den ausgehenden Verkehr. Dabei beeinflussen die Schwankungen des über die Peer-to-Peer-Verbindungen empfangenen Verkehrs ausschließlich den ausgehenden Verkehr zu den Kunden-NGN. Wohingegen der von den Kunden-NGN erhaltene Verkehr seine Änderungen im Fehlerfall teilweise an den ausgehenden Verkehr der Peer-to-Peer-Verbindungen weitergibt. Trotzdem unterscheiden sich die Verkehrsschwankungen bei den einzelnen Verbindungen, wobei die Quantile bei Betrachtung aller Fehler bei NGN3 4,0% nicht übersteigen.

4.4.2.5 Dimensionierungsregeln für das KGID

Aus den in den letzten Abschnitten ermittelten Ergebnissen werden im folgenden Regeln für die Überdimensionierung der Interdomain-Verbindungen für das KGID abgeleitet. Mit dieser Maßnahme soll sichergestellt werden, daß bei einer Interdomain-Topologieänderung genügend Kapazität für den umgelegten Verkehr zur Verfügung steht.

Für die Stub-NGN lassen sich relativ einfach die benötigten Kapazitäten für die beiden Interdomain-Verbindungen finden. Der Hauptgrund für die zweite Verbindung eines Stub-NGN besteht darin, daß bei einem Ausfall einer der beiden Verbindungen noch eine Ersatzverbindung zur Verfügung steht. Um diese sinnvoll nutzen zu können, müssen beide Interdomain-Verbindungen so ausgelegt werden, daß jeweils eine Verbindung den kompletten Interdomain-Verkehr des Stub-NGN transportieren kann. Mit dieser Dimensionierung kann ein Stub-NGN alle Verkehrsverschiebungen aufgrund von Interdomain-Topologieänderungen bewältigen. Fällt die zweite Verbindung ebenfalls aus, existiert keine Verbindung zu den restlichen NGN mehr, und somit ist diese Dimensionierung für alle denkbaren Fehler ausreichend.

Solch eine einfache Regel läßt sich für die anderen NGN nicht finden, da sich hierbei viele verschobene Verkehrsflüsse überlagern und eine Dimensionierung für alle Fehler viel zu hohe Kapazitätsanforderungen stellen würde. Deshalb ist es sinnvoll, nur für die Mehrzahl an Fehlern ausreichend Kapazität anzubieten. Generell gilt, daß je nachdem welche Fehlerbilder man bei der Dimensionierung berücksichtigt, man unterschiedliche Kapazitätsanforderungen erhält.

Zieht man nur die Schwankungen gemittelt über alle Verbindungen und Fehlerbilder in Betracht, so sind die Anforderungen an die zusätzliche Kapazität sehr gering. Werden dagegen die Verkehrsschwankungen auf einer einzelnen Verbindung bei alle direkten Fehlern für die Dimensionierung herangezogen, ist wesentlich mehr Kapazität erforderlich.

Die Dimensionierungsvorschriften für das KGID nutzen einen Mittelweg. Hierbei werden alle Fehlerbilder und die Verkehrsverschiebungen jeder einzelnen Verbindung sortiert nach deren Kategorie für die Bestimmung der notwendigen Kapazität herangezogen. Außerdem werden nur die Verkehrsschwankungen für 99,5% aller Fehler berücksichtigt, was die großen und sehr selten auftretenden Änderungen des Verkehrs bei der Dimensionierung ausschließt. Damit ermöglichen die folgenden Vorschriften eine kapazitätseffiziente Dimensionierung der Interdomain-Verbindungen.

Bei der Aufstellung der Dimensionierungsregeln für die kleinen Tier2-NGN ist es vorteilhaft, die Vorschriften für die Stub-NGN mit einzubeziehen. Ein Großteil der Verbindungen der kleinen Tier2-NGN zu den Kunden-NGN sind Verbindungen zu Stub-NGN. Werden diese wie vorgeschlagen ausgelegt, müssen nur die Verkehrsverschiebungen des von kleinen Tier2-NGN erzeugten bzw. terminierten Verkehrs für die Dimensionierung betrachtet werden.

Wie die Ergebnisse von NGN82 und NGN213 zeigen, ist eine Überdimensionierung der Verbindungen zu den Betreiber-NGN mit 30% ausreichend. Allerdings weisen die Änderungen bei den Verbindungen zu den Kunden-NGN bei NGN82 noch auf ein anderes Problem hin. Besitzt ein NGN nur zwei Verbindungen zu seinen Betreibern und die restlichen verbinden das NGN mit seinen Kunden, sollten die beiden Verbindungen mit den Betreiber-NGN so ausgelegt werden, daß beide Verbindungen jeweils den kompletten Verkehr allein übertragen können.

Werden diese Ausnahmeregeln für die NGN mit zwei Verbindungen zu deren Betreiber-NGN beachtet, reichen 30% mehr Kapazität für die Interdomain-Verbindungen der großen Tier2-NGN zu deren Betreiber- und Kunden-NGN ebenfalls aus. Für die Peer-to-Peer-Verbindungen müssen aber nur 5% an zusätzlicher Kapazität bereitgestellt werden.

Dieselbe Menge an extra Kapazität wird auch bei den Transit-NGN für die Peer-to-Peer-Verbindungen benötigt. Aufgrund der größeren Unterschiede bei den Verbindungen eines Transit-NGN zu seinen Kunden lohnt sich eine Unterteilung der Verbindungsdimensionierung anhand der Größe der Nachbar-NGN. Die Verbindungen zu den großen Tier2-NGN sollten mit 30% mehr Kapazität ausgestattet werden. Für die anderen Kunden-Verbindungen ist eine Überdimensionierung um maximal 10% ausreichend.

Mittels der in diesem Abschnitt erstellten Dimensionierungsregeln ist sichergestellt, daß für mehr als 99,5% aller Fehlerfälle genügend Kapazität für die daraus resultierenden Verkehrsverschiebungen zur Verfügung steht. Die meiste Kapazität muß für die NGN, die nur zwei Verbindungen zu ihren Betreiber-NGN besitzen, aufgebracht werden. Bei den restlichen NGN müssen maximal 30% an zusätzlicher Kapazität bereitgestellt werden, um auch im Fall von Verkehrsverschiebungen verursacht durch Interdomain-Topologieänderungen die Dienstqualität sicherzustellen.

4.4.3 Auswirkungen auf die Netzzugangskontrollverfahren

Nachdem man die Interdomain-Verbindungen so dimensioniert, daß für den Großteil der Interdomain-Topologieänderungen ausreichend Kapazität für die Verschiebung von QoS-Verkehr zur Verfügung steht, muß noch ein weiteres Problem gelöst werden.

In Abschnitt 4.3 wurde die BGP+-Reaktion auf eine Topologieänderung so definiert, daß der Interdomain-QoS-Verkehr nach der Routing-Änderung ohne eine sofortige Anpassung der Ressourcenreservierungen auf den neuen Pfad umgelegt wird. Damit erhalten die betroffenen NGN QoS-Verkehr, über den sie im voraus keine Informationen erhalten haben und somit auch keine Ressourcen explizit belegen konnten.

Aber ein NGN muß sicherstellen, daß es QoS-Verkehr ohne Beeinträchtigung der Dienstqualität weiterleiten kann. Dazu gehört auch, daß kein NGN mehr QoS-Verkehr verschickt als Ressourcen dafür reserviert wurden. Für diesen Zweck muß ein NGN den eingehenden QoS-Verkehr überwachen und kontrollieren, ob er die zugesagten Randbedingungen erfüllt. Die dazu eingesetzten Mechanismen werden als Netzzugangskontrollverfahren bezeichnet.

Erhält ein NGN zusätzlichen QoS-Verkehr aufgrund einer Routing-Änderung, kann das Netzzugangskontrollverfahren je nach Konfiguration diesen QoS-Verkehr verwerfen oder die Dienstqualität einschränken oder sogar verweigern. Um dies zu vermeiden, muß das Netzzugangskontrollverfahren einen Mechanismus enthalten, der für eine gewisse Zeit QoS-Verkehr auch ohne gültige Reservierung in das NGN hinein- und wieder herausläßt.

Die Lösung dieses Problems besteht in der Einführung einer neuen QoS-Klasse. Für die Sicherstellung der Dienstqualität, die schon wegen der Routing-Änderung beeinträchtigt ist, muß die neue QoS-Klasse die höchste Klasse für QoS-Nutzverkehr sein. Diese Klasse darf nur von QoS-Verkehr, der von Routing-Änderungen betroffen ist, benutzt werden, um den besonderen Bedingungen gerecht werden zu können.

Da das QoS-NSLP nach der Routing-Änderung versuchen wird, die Ressourcenreservierungen des betroffenen QoS-Verkehrs zu aktualisieren, wird der verschobene QoS-Verkehr, für den keine Reservierung existiert, mit der Zeit abnehmen. Damit wird die höchste QoS-Klasse nur kurzzeitig in Anspruch genommen, und die Verkehrscharakteristik entspricht einer Spitze zu Beginn, die über die Zeit abfällt. Damit das Netzzugangskontrollverfahren für diese Art von Verkehr, wie sie in [**CSG+06**] beschrieben wird, konfiguriert werden kann, müssen im SLA zwischen den Netzbetreibern der NGN die Größe dieser Spitze und die maximale Zeitdauer für das Senden des QoS-Verkehrs mit der höchsten Priorität festgelegt werden.

Desweiteren muß die QoS-Klasse für den umgelegten QoS-Verkehr global gültig sein, und das NGN, das den QoS-Verkehr umlegt, muß die alte Markierung der Datenpakete durch die der höchsten QoS-Klasse beim Verlassen des NGN ersetzen.

4.5 Zusammenfassung

Ausgangspunkt der Betrachtungen in diesem Kapitel ist die Evolution des heutigen Internets zu einem Netz, das die Dienstqualität für die Datenübertragung zwischen mehreren NGN zusagen kann. Da diese Entwicklung eine gewisse Zeit in Anspruch nehmen wird, kommt es am Anfang zu einer Inselbildung, wobei sich diese Inseln mit der Zeit vergrößern und zusammenwachsen werden. Währenddessen kann nur innerhalb dieser Inseln Verkehr transportiert werden, für den die Dienstqualität garantiert werden muß. Die sich aus diesem Prozeß ergebenden Anforderungen wurden aufgestellt und die existierenden Lösungsansätze daran gemessen. Keiner der existierenden Mechanismen konnte die gestellten Rahmenbedingungen vollständig erfüllen. Aus diesem Grund wurde das Konzept für Garantierte Interdomain-Dienstqualität (KGID) entwickelt.

Ein Bestandteil des KGID ist das in Kapitel 3 entworfene BGP+, das an die erweiterten Anforderungen der Interdomain-Übertragung von QoS-Verkehr angepaßt wurde. Wegen der Unkenntnis der verfügbaren Interdomain-Ressourcen und der Nichtvorhersagbarkeit des BGP+-Routing kann die Dienstqualität für den QoS-Verkehr nach einer Interdomain-Topologieänderung nur aufrecht erhalten werden, wenn dafür ausreichend Ressourcen im voraus zur Verfügung gestellt werden.

Dabei galt es herauszufinden, wieviel zusätzliche Ressourcen für diese Interdomain-Verkehrsverschiebungen notwendig sind. Für diesen Zweck wurde ein Modell entworfen, anhand dessen die Schwankungen des Interdomain-Verkehrs beim Ausfall einer Interdomain-Verbindung untersucht wurden. Basierend auf den erzielten Ergebnissen wurden Dimensionierungsregeln für die Interdomain-Verbindungen aufgestellt, so daß in mehr als 99,5% aller Einzelverbindungsausfälle jeder Interdomain-Verbindung ausreichend Kapazität für die Interdomain-Verkehrsverschiebungen zur Verfügung steht. Für die NGN (Stub- und kleine Tier2-NGN), die nur zwei Verbindungen zu ihren Betreiber-NGN besitzen, müssen beide Verbindungen so ausgelegt werden, daß jeweils eine den kompletten Interdomain-QoS-Verkehr tragen kann. Die Peer-to-Peer-Verbindungen erfordern nur eine Überdimensionierung von 5%. Bei den Interdomain-Verbindungen von den Transit-NGN zu den kleinen Tier2-NGN reichen 10% mehr Kapazität aus. Alle anderen Interdomain-Verbindungen benötigen 30% mehr, um ausreichend Kapazität für die Interdomain-Verkehrsverschiebungen bereitzuhalten.

Damit der QoS-Verkehr die bereitgestellte Kapazität nach einer Interdomain-Topologieänderung auch benutzen kann, muß das in einem NGN möglicherweise eingesetzte Netzzugangskontrollverfahren an diese Verkehrsverschiebungen ohne vorherige Ressourcenreservierung angepaßt werden. Für diesen Zweck wird eine weitere, allen NGN bekannte und global eindeutige QoS-Klasse eingeführt. Diese muß die höchste Priorität aller Interdomain-QoS-Klassen für den Datentransport erhalten und kann nur kurzzeitig nach einer Interdomain-Topologieänderung verwendet werden.

Aus diesen Mechanismen und Vorschriften wurde ein Konzept entwickelt, das die Interdomain-Dienstqualität auch nach einem Verbindungsausfall garantieren kann. Welche Konsequenzen diese Interdomain-Datenübertragung von QoS-Verkehr auf die Netzdimensionierung eines NGN innerhalb einer QoS-Insel hat, wird im folgenden Kapitel ausführlich diskutiert.

Kapitel 5

Intradomain-Netzdimensionierung für QoS-Verkehr

In Abschnitt 4.3.3.2 wurde erläutert, daß bei der Netzdimensionierung zwei Fälle berücksichtigt werden müssen, um zwischen mehreren NGN die Dienstqualität zu garantieren. Zum einen muß auf den Interdomain-Verbindungen genügend Kapazität für die Interdomain-Verkehrsverschiebungen bereitgestellt werden, und zum zweiten muß ein NGN gewährleisten, daß es ohne Beeinträchtigung der Dienstqualität den QoS-Verkehr durch das NGN transportieren und zu seinem Ziel weiterleiten kann.

Für die erste Forderung wurden im letzten Kapitel die entsprechenden Regeln für die Dimensionierung der Interdomain-Verbindungen aufgestellt. Der zweite Punkt wird im folgenden untersucht und dafür ein neues Netzdimensionierungsverfahren erarbeitet.

5.1 Transport von Interdomain-QoS-Verkehr

Wird zwischen verschiedenen NGN innerhalb einer QoS-Insel QoS-Verkehr ausgetauscht, vertrauen alle Netzbetreiber dieser NGN darauf, daß die anderen NGN die geforderte Dienstqualität zusagen und einhalten können. Bei der dabei entstehenden Kette des Vertrauens der beteiligten NGN muß jedes NGN diesem Vertrauen gerecht werden. Dazu gehört nicht nur, die Rahmenbedingungen aus Abschnitt 4.1 zu erfüllen, sowie genügend Kapazität für die Interdomain-Topologieänderungen, wie in Abschnitt 4.4 beschrieben, bereitzustellen, sondern auch ein NGN so zu dimensionieren, daß es den Interdomain-QoS-Verkehr mit der zugesicherten Dienstqualität übertragen kann.

Das beinhaltet, daß der von dem NGN erzeugte Interdomain-QoS-Verkehr das NGN verlassen, der in dem NGN terminierte Interdomain-QoS-Verkehr in das NGN gelangen sowie der Transit-QoS-Verkehr das NGN passieren kann. Um welchen und um wieviel Interdomain-QoS-Verkehr es sich dabei handelt, wird in den SLA zwischen den einzelnen NGN vertraglich festgeschrieben.

Damit besteht die Aufgabe bei der Netzplanung, diesen Anforderungen bei der Dimensionierung eines NGN gerecht zu werden.

Der wesentliche Unterschied zu der bekannten Netzplanung für BE-Verkehr besteht darin, daß der Netzbetreiber eines NGN die Zusage erteilt, den QoS-Verkehr zu dessen Ziel zu transportieren und dabei unter Umständen auf die Unterstützung der anderen NGN angewiesen ist. Geht man davon aus, daß die betroffenen NGN ihre im SLA zugesagten Leistungen erbringen, besteht das Problem, daß das NGN seinen Kunden den Transport von QoS-Verkehr garantiert, unabhängig davon, zu welchem Ziel dieser Verkehr weitergeleitet werden soll. Es besteht die Möglichkeit, daß der gesamte Interdomain-QoS-Verkehr das NGN an einem Ausgangs-Router verläßt. Da die Kunden nicht daran interessiert sind, die Ziele ihres Interdomain-QoS-Verkehr festzulegen und wahrscheinlich dies auch nicht voraussagen können, muß dies bei der Dimensionierung des NGN berücksichtigt werden.

Wie in Abschnitt 4.4 erläutert wurde, kann auch eine Interdomain-Routing-Änderung dazu führen, daß sich der Ausgangs-Router für den Interdomain-QoS-Verkehr ändert. Daraus resultiert ebenfalls die Anforderung, ein NGN so auszulegen, daß der Interdomain-QoS-Verkehr unabhängig von dem Router ist, über den dieser Verkehr das NGN verläßt.

Welche existierenden Lösungen dafür in Frage kommen bzw. mit dieser Aufgabenstellung in Zusammenhang stehen, wird in den folgenden Abschnitten behandelt. Dabei werden diese Ansätze an ihrer Robustheit gegenüber dem Wechsel des Ausgangs-Routers bewertet. Ein robuster Mechanismus ermöglicht es, daß der ein NGN verlassende Interdomain-QoS-Verkehr ohne Beeinträchtigung der Dienstqualität über alle möglichen Ausgangs-Router dieses NGN versendet werden kann.

5.1.1 Nicht robuste Netzdimensionierung

Bei der nicht robusten Netzdimensionierung, wie sie in Abschnitt 2.3 beschrieben wurde, benutzt man eine Verkehrsmatrix, wo die Verkehrsflüsse zwischen zwei Knoten genau definiert sind. Somit wird die Quelle und die Senke sowie der Verkehrsfluß zwischen diesen beiden genau festgelegt. In einigen Untersuchungen zur Robustheit gegenüber Verkehrsschwankungen variieren diese Verkehrsflüsse um einen gewissen Anteil, wie z.B. in [VCP+03] Änderungen von 10% betrachtet werden.

Dieser Ansatz mit der statischen Festlegung der Quellen und Ziele paßt nicht zu den Anforderungen der Netzdimensionierung für Interdomain-QoS-Verkehr, wo das Ziel bzw. der Ausgangs-Router variabel sind. Damit ist eine andere Lösung notwendig.

5.1.2 Robuste Netzdimensionierung

Das Problem der Netzdimensionierung für Verkehr mit unbekanntem Ziel tritt ebenfalls bei der Planung von Netzen für *Virtual Privat Networks* (VPN) auf [GLH+00]. Dafür existieren viele verschiedene Lösungsvorschläge, wie z.B. [BAK05, DGG+02, KRS+01, IRY02]. Die meisten

dieser Ansätze bauen auf dem sogenannten Schlauchmodell auf, was zuerst in [DGG+99] vorgestellt wurde. Es geht davon aus, daß der Verkehr aggregiert zu einem Schlauch in das Netz hineinfließt und die Verkehrsmenge dieses Schlauchs zu allen Ausgangs-Routern weitergeleitet wird, was in Abbildung 5.1 veranschaulicht wird.

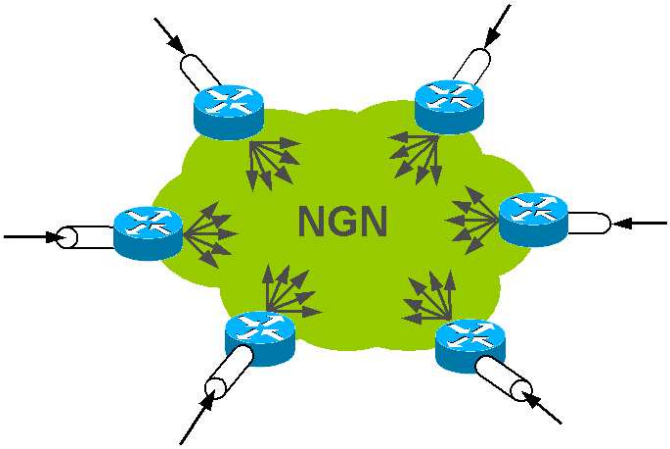


Abbildung 5.1: Das Schlauchmodell

Aufgrund der unbeschränkten Zielauswahl stellt dieser Ansatz das andere Extrem zu der Definition einer statischen Verkehrsmatrix dar. Allerdings hat diese Lösung einen großen Kapazitätsbedarf für die einzelnen Verbindungen, da das Netz so dimensioniert wird, daß der gesamte eingehende Verkehr komplett zu allen anderen Ausgangs-Routern weitergeleitet werden kann. Aufgrund dieser Auslegung bietet diese Modellierung die maximale Robustheit und wird deswegen im folgenden als robuste Netzdimensionierung bezeichnet.

5.1.3 Netzdimensionierung unter Verwendung von Netzzugangskontrollverfahren

An den Routern am Rand eines NGN können Netzzugangskontrollverfahren einsetzen werden, um damit die Menge des eingehenden QoS-Verkehrs zu überwachen und sicherzustellen, daß die Verkehrsmenge, die im SLA festgeschriebene Grenze bzw. die Schwelle, bis zu der das NGN die Dienstqualität garantieren kann, nicht überschritten wird. Handelt es sich bei dem überwachten QoS-Verkehr um verbindungsorientierten Verkehr, z.B. VoIP, bezeichnet man das Netzzugangskontrollverfahren als *Call-Admission Control* (CAC), und wenn der QoS-Verkehr nicht näher spezifiziert ist, spricht man von *Network Admission Control* (NAC).

Beide Mechanismen unterscheiden sich anhand der Algorithmen und der dabei verwendeten Parameter, mit denen sie entscheiden, ob eine Ressourcenreservierung akzeptiert wird oder nicht. Ansonsten können beide gezielt Datenpakete eines Verkehrsflusses löschen, wenn dieser eine vorher definierte Schwelle überschreitet.

Für die Verwendung einer CAC bzw. einer NAC in einem Netz gibt es die verschiedensten Ansätze. Ein sehr guter Überblick und ein Vergleich der gängigen Architekturen wird in [Rie04] und [Men04] geboten. Je nachdem, welche Architekturen eingesetzt werden, ergeben sich die verschiedensten Anforderungen an die Dimensionierung eines Netzes. Wählt man die einfachste Konfiguration, wo die CAC oder NAC nur an den Eingangs-Routern betrieben wird und alle nachfolgenden Router nicht mit berücksichtigt, entsteht das Problem, daß Verkehr unabhängig vom Ziel in das Netz hineingelassen wird. Damit muß solch ein Netz nach dem Schlauchmodell dimensioniert werden. Aber die daraus resultierenden großen Kapazitätsanforderungen dieses Ansatzes führen in vielen Untersuchungen dazu, daß ein anderes Konzept ausgewählt wird, wie z.B. in [RBF03].

Prinzipiell hat die Netzdimensionierung unter Berücksichtigung eines Netzzugangskontrollverfahrens andere Randbedingungen als die Netzdimensionierung für Interdomain-QoS-Verkehr. Bei der Netzdimensionierung eines Netzes, das ein Netzzugangskontrollverfahren einsetzt, bestimmt das verwendete Netzzugangskontrollverfahren die Punkte im Netz, an denen eine obere Grenze an Ressourcen nicht überschritten werden darf. Die Aufgabe der Netzdimensionierung besteht darin, das Netz so auszulegen, daß für den durch diese Grenzwerte festgelegte Maximalverkehr ausreichend Ressourcen zur Verfügung stehen.

Soll ein Netz für Interdomain-QoS-Verkehr dimensioniert werden, ist die einzige feste Größe der an einem Eingangs-Router eintreffende Interdomain-QoS-Verkehr. Damit entspricht dieses Problem der einfachsten Konfiguration, wo nur an den Eingangs-Routern eine CAC oder NAC eingesetzt wird.

5.1.4 Statistische Netzdimensionierung

Trotz der großen Kapazitätsanforderungen der CAC-Architektur, wo nur am Eingangs-Router überprüft wird, ob der empfangene QoS-Verkehr die vordefinierte Schwelle nicht überschreitet, wird in Kapitel 5 von [Sha05b] diese CAC-Architektur wegen ihrer Einfachheit favorisiert. Aufgrund dieser Wahl findet das Schlauchmodell bei der Netzdimensionierung Anwendung und dies führt zu hohen Kapazitätsanforderungen. Um diese zu reduzieren, wird die als statistische Netzdimensionierung bezeichnete Methode in [Sha05b] vorgeschlagen.

Die Idee der statistischen Netzdimensionierung ist die Einführung einer Wahrscheinlichkeit, mit der die Dienstqualität der übertragenen Daten beeinträchtigt wird. Dazu werden alle möglichen Ausgangs-Router betrachtet, zu denen der QoS-Verkehr weitergeleitet werden kann, und die neu eingeführte Wahrscheinlichkeit gibt an, in wieviel Prozent aller Möglichkeiten die im Netz bereitgestellte Kapazität nicht ausreicht, um den kompletten Verkehr zu transportieren. Somit bestimmt die Wahl dieser Wahrscheinlichkeit, wie groß der Kapazitätsbedarf für die Übertragung des QoS-Verkehrs ist. Abbildung 5.2 aus [Sha05b] zeigt den Zusammenhang zwischen der benötigten Kapazität und dem Wert dieser Wahrscheinlichkeit anhand eines Beispielnetzes.

Wird die Wahrscheinlichkeit ε auf Null gesetzt, ist der Kapazitätsbedarf des Netzes gleich der Dimensionierung basierend auf dem Schlauchmodell (in [Sha05b] als *strictly robust* bezeichnet).

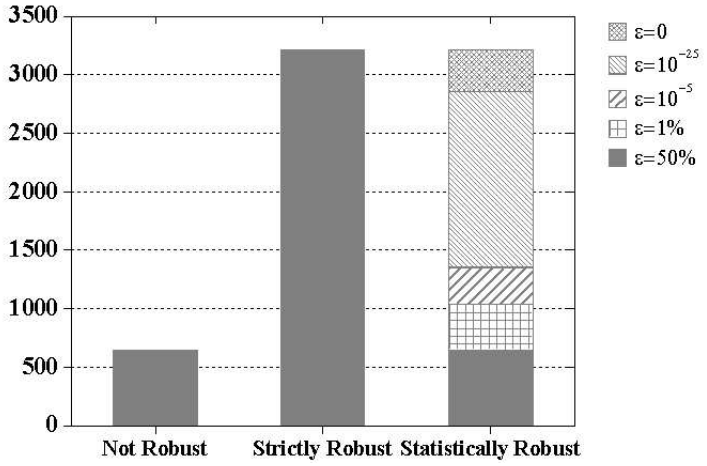


Abbildung 5.2: Kapazitätsbedarf der statistischen Dimensionierung [Sha05b]

Beträgt $\varepsilon = 50\%$, dann sinkt der Kapazitätsbedarf auf das Niveau der nicht robusten Netzdimensionierung (in Abbildung 5.2 *not robust* genannt). Allerdings ist bei dieser starken Beeinträchtigung der Dienstqualität keine zufriedenstellende Übertragung des QoS-Verkehrs mehr möglich. Ist die Wahrscheinlichkeit ε nur minimal größer als Null, können schon beachtliche Einsparungen der im Netz benötigten Kapazität erzielt werden, da dabei die Extremfälle mit dem größten Kapazitätsbedarf vernachlässigt werden.

Somit bietet die statistische Netzdimensionierung eine Möglichkeit, die hohen Kapazitätsanforderungen des Schlauchmodells zu reduzieren. Allerdings muß ein Netzbetreiber entscheiden, ob ihm die Verringerung der benötigten Kapazität die daraus resultierende Reduzierung der Robustheit wert ist.

5.2 Dimensionierung mit Rücksicht auf das BGP+-Routing

In diesem Abschnitt wird eine neue Methode für die Netzdimensionierung beschrieben, die den Kapazitätsbedarf für den Interdomain-QoS-Verkehr eines NGN ohne Abstriche bei der Dienstqualität reduziert. Da das Interdomain-Routing mit BGP+ festlegt, über welchen Eingangs- bzw. Ausgangs-Router der Interdomain-QoS-Verkehr das NGN betritt bzw. verläßt, kann man mit dem Wissen über die Routing-Entscheidungen von BGP+ die Anzahl der möglichen Ausgangs-Router reduzieren. Diese Einschränkung der Auswahlmöglichkeiten führt zu Einsparungen bei den Kapazitätsanforderungen. Im Gegensatz zur statistischen Netzdimensionierung wird die Robustheit dadurch nicht reduziert, da der Interdomain-QoS-Verkehr ein NGN nur über die Ausgangs-Router verlassen kann, die das Interdomain-Routing mit BGP+ zuläßt. Die Robustheit wird aber nur dann nicht beeinträchtigt, wenn alle aus Sicht von BGP+ zulässigen Ausgangs-Router berücksichtigt werden. Da diese Robustheit nur für den Interdomain-QoS-Verkehr erforderlich ist, wird im folgenden die Einteilung des Verkehrs und eine Meßmethode beschrieben, mit der der Intra- und Interdomain-QoS-Verkehr unterschieden werden kann.

5.2.1 Einteilung des Verkehrs

Grundsätzlich kennt ein NGN vier verschiedenen Arten von QoS-Verkehr. Zum einen ist dies der Verkehr, der das NGN nicht verläßt. Das heißt, daß die Quelle und die Senke innerhalb desselben NGN liegen. Bei den restlichen drei Arten liegen entweder die Quelle, die Senke oder alle beide außerhalb des NGN.

Wenn der QoS-Verkehr nur innerhalb des NGN weitergeleitet wird, tritt das Problem des unbekannten Ausgangs-Routers nicht auf. Deswegen kann das Netz für diesen Intradomain-QoS-Verkehr mittels der nicht robusten, aber kapazitätseffizienten Netzdimensionierung (s. Abschnitt 5.1.1) ausgelegt werden.

Erst bei den anderen drei Verkehrsarten ist der Ausgangs- bzw. Eingangs-Router des QoS-Verkehrs aufgrund des Interdomain-Routing variabel. Dabei handelt es sich um Interdomain-QoS-Verkehr, der in einem NGN erzeugt und zu anderen NGN weitergeleitet wird, der von anderen NGN empfangene QoS-Verkehr, der in einem NGN terminiert wird, sowie um Transit-QoS-Verkehr, der das NGN nur passiert. Für diese drei Arten von QoS-Verkehr muß das Schlauchmodell angewendet werden. Der als Schlauch zusammengefaßte QoS-Verkehr ist damit entweder der von benachbarten NGN empfangene QoS-Verkehr oder/und der an einem Knoten des NGN erzeugte QoS-Verkehr. Über welchen Ausgangs-Router dieser QoS-Verkehr an das dementsprechende benachbarte NGN übergegeben wird, legt das Interdomain-Routing-Protokoll fest. Das wird genutzt, um mit der Kenntnis des Routing von BGP+ die möglichen Ausgangs-Router für die jeweiligen Schläuche einzuschränken und ein detaillierteres Schlauchmodell daraus zu entwickeln [**Sch06b**].

5.2.2 Ermitteln der Verkehrseinteilung

Um dieses detailliertere Schlauchmodell anwenden zu können, müssen die vier Arten von QoS-Verkehr innerhalb eines NGN ermittelt werden. Da die Verkehrsflüsse eines schon bestehenden Netzes als Basis für die Vorhersage der Verkehrsmatrix genutzt werden können, kann die Messung dieser Verkehrsflüsse auch für die Bestimmung der Verkehrseinteilung genutzt werden. Am besten dafür geeignet ist die in [FGL+01] vorgestellte Meßmethode, da diese genau dafür entwickelt wurde, um die Unterteilung des Verkehrs in die vier Gruppen meßtechnisch zu ermitteln. Dabei wurde von einer Netzstruktur, wie sie in Abbildung 5.3 zu sehen ist, ausgegangen.

Im Gegensatz zu dieser Arbeit unterscheidet sich der Aufbau eines autonomen Systems aus Abbildung 5.3 etwas, aber er läßt sich auf den hier verwendeten Aufbau übertragen. In [FGL+01] wird angenommen, daß ein autonomes System nicht nur aus dem sogenannten Kernnetz (die *backbone links* in Abbildung 5.3) besteht, sondern es auch noch Zugangsnetze (die *access links* in Abbildung 5.3) besitzt. Diese Zugangsnetze werden in dieser Arbeit nicht betrachtet. Aber sie liefern ihren aggregierten Verkehr an die Knoten des Kernnetzes, und damit kann dieser Verkehr auch als in diesen Knoten generierter Verkehr verstanden werden. Aufgrund der Unterscheidung zwischen Zugangs- und Kernnetz verlassen auch alle Verkehrsflüsse das in Abbildung 5.3 dargestellte Kernnetz.

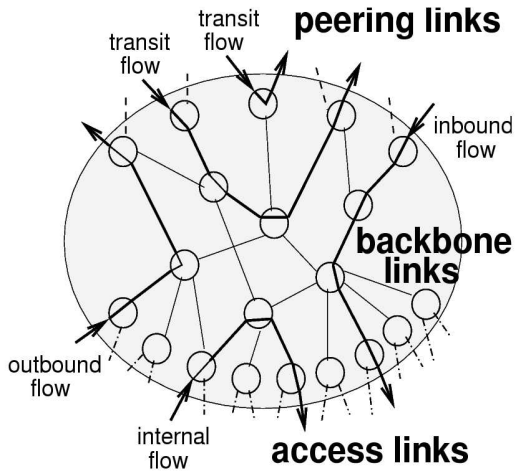


Abbildung 5.3: Charakterisierung der Verkehrsflüsse [FGL+01]

Die als *transit flow* bezeichneten Verkehrsflüsse werden in dieser Arbeit als Transit-QoS-Verkehr bezeichnet. Wohingegen der von einem NGN empfangene Interdomain-QoS-Verkehr dem *inbound flow* entspricht. Das Gegenstück dazu, der *outbound flow* ist gleich dem von einem NGN versendeten Interdomain-QoS-Verkehr. Der QoS-Verkehr, der ein NGN nicht verläßt, kann dem sogenannten *internal flow* gleichgesetzt werden.

Mit dieser Zuordnung kann der in [FGL+01] vorgestellte Ansatz dazu verwendet werden, die Anteile der vier verschiedenen Verkehrsarten zu bestimmen. Allerdings muß diese Methode noch um die Unterscheidung zwischen BE- und QoS-Verkehr erweitert werden. Das heißt, daß bei der Ermittlung der Werte die Markierung der IP-Pakete mit in die Auswertung aufgenommen werden muß.

Kann bei der Netzdimensionierung nicht auf die Daten eines bestehenden Netzes zurückgegriffen werden, müssen Annahmen über die Aufteilung des Verkehrs getroffen werden. Dabei können die in den SLA mit anderen Netzbetreibern festgelegten Verkehrsmengen einen Anhaltspunkt geben.

5.2.3 Einschränkung der möglichen Ausgangs-Router

Nachdem erläutert wurde, auf welche Teile des QoS-Verkehrs das Schlauchmodell angewendet werden muß und wie man diese Teile ermittelt, wird im folgenden beschrieben, wie man die Kenntnisse über das Interdomain-Routing für die Einschränkung der möglichen Ausgangs-Router nutzen kann.

Da jeder Netzbetreiber die Bestimmung der besten Pfade von BGP+ mittels der Policies individuell gestalten kann und die dafür benutzten Regeln nicht öffentlich verfügbar sind, werden nachfolgend nur die in [Gao01] definierten verallgemeinerten Policies verwendet. Für deren Anwendung ist es notwendig, die schon in Kapitel 4 benutzten Unterscheidungen der NGN anhand

ihrer Größe sowie der Interdomain-Verbindungen aufgrund der Geschäftsbeziehungen vorzunehmen.

Resultierend aus diesen Einteilungen können die verallgemeinerten Policies (s. Tabelle 2.1) genutzt werden, um nicht alle möglichen Ausgangs-Router bei der Bestimmung der Verkehrsflüsse basierend auf dem Schlauchmodell berücksichtigen zu müssen. In Abhängigkeit der Art des NGN – Transit-, Tier2- und Stub-NGN – ergibt sich ein detaillierteres Schlauchmodell, wo nicht mehr jeder Interdomain-QoS-Verkehrsfluß zu allen Ausgangs-Routern weitergeleitet wird.

Bei den Transit-NGN ergeben sich aus den verallgemeinerten Policies zwei Regeln, die beachtet werden müssen. Der Interdomain-QoS-Verkehr, den ein Transit-NGN von seinen Kunden-NGN empfängt, wird entweder zu einem über eine Peer-to-Peer-Verbindung angeschlossenes NGN (symbolisiert durch die gestrichelten Pfeile in Abbildung 5.4) oder an ein anderes Kunden-NGN weitergeleitet (in Abbildung 5.4 als durchgehende Pfeile markiert).

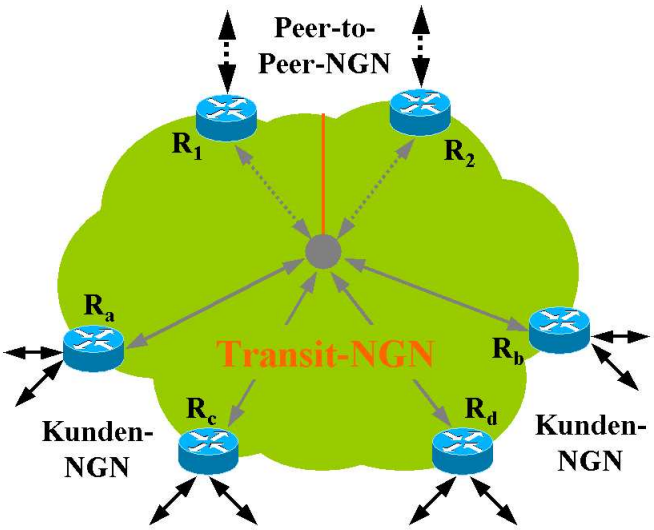


Abbildung 5.4: Verkehrsflüsse durch ein Transit-NGN

Im folgenden wir ein NGN, das mittels einer Peer-to-Peer-Verbindung erreichbar ist, als Peer-to-Peer-NGN bezeichnet. Im Fall eines Transit-NGN sind dessen Peer-to-Peer-NGN meistens ebenfalls Transit-NGN. Empfängt ein Transit-NGN von einem Peer-to-Peer-NGN Verkehr, kann dieser nur an ein Kunden-NGN weitergegeben werden und nicht an andere Peer-to-Peer-NGN (orangefarbene Trennlinie in Abbildung 5.4).

Betrachtet man ein Tier2-NGN, kann der von einem Kunden-NGN empfangene Interdomain-QoS-Verkehr an andere Kunden-NGN, an ein Betreiber-NGN oder an ein Peer-to-Peer-NGN weitergeleitet werden. Erhält ein Tier2-NGN Interdomain-QoS-Verkehr von einem Betreiber-NGN oder von einem Peer-to-Peer-NGN, kann dieser Interdomain-QoS-Verkehr nur an ein Kunden-NGN übergeben werden, was mittels orangefarbener Trennlinien in Abbildung 5.5 symbolisiert ist.

Für den von einem Tier2-NGN erzeugten Interdomain-QoS-Verkehr gelten dieselben Regeln wie

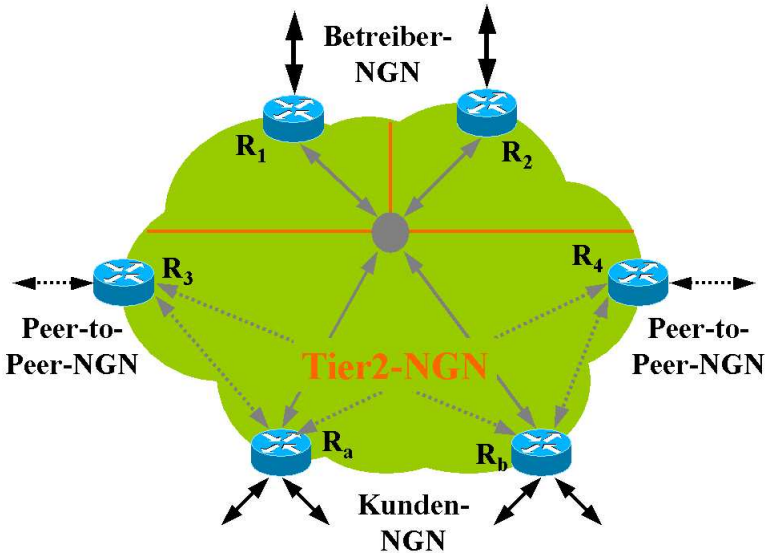


Abbildung 5.5: Verkehrsflüsse durch ein Tier2-NGN

für den Interdomain-QoS-Verkehr, der von den Kunden-NGN empfangen wird, das bedeutet, daß dieser Verkehr auf alle drei Verbindungsarten aufgeteilt wird. Ebenfalls über alle drei Verbindungstypen kann ein Tier2-NGN Verkehr empfangen, der in diesem NGN terminiert wird.

Bei den Stub-NGN kann der erzeugte Interdomain-QoS-Verkehr nur über die beiden vorhandenen Interdomain-Verbindungen an deren Betreiber-NGN weitergegeben werden. Genauso kann der in den Stub-NGN terminierte Verkehr nur über die beiden Verbindungen in das NGN gelangen, wie es in Abbildung 5.6 dargestellt ist.

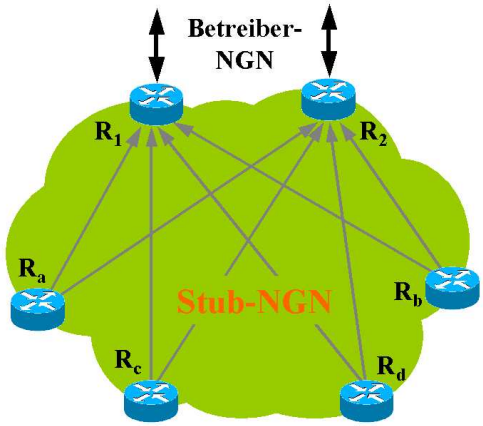


Abbildung 5.6: Verkehrsflüsse durch ein Stub-NGN

Allein mit diesen einfachen Regeln, basierend auf den verallgemeinerten Policies, läßt sich die Auswahl der Ausgangs-Router einschränken, womit eine Reduzierung der Kapazität für den Transport von Interdomain-QoS-Verkehr in einem NGN zu erwarten ist. Im allgemeinen kennt

der Netzbetreiber die in seinem NGN eingesetzten Policies, und mit diesem wesentlich detaillierteren Wissen kann die Vorauswahl der Ausgangs-Router weiter eingeschränkt werden, sodaß nochmals bessere Ergebnisse erzielt werden können.

5.3 Dimensionierung eines NGN

Nachdem in den letzten Abschnitten die Idee erklärt wurde, wie durch die Berücksichtigung des Routing mit BGP+ und den dazugehörigen Policies das Schlauchmodell angepaßt werden kann, wird im folgenden anhand eines Beispiels gezeigt, welche Einsparungen im Rahmen der Annahmen möglich sind.

5.3.1 Untersuchte Szenarien

Als Basis für den Vergleich der verschiedenen Dimensionierungsmethoden wurde das deutsche Netz des NOBEL-Projekts ausgewählt, wie es als NOBEL-Germany-Network in [Zus06] beschrieben ist. Dieses Netz besteht aus siebzehn Knoten und sechsundzwanzig bidirektionalen Verbindungen und ist in Abbildung 2.5 dargestellt.

Desweiteren ist in [Zus06] auch eine dazugehörige Verkehrsmatrix mit 121 symmetrischen Verkehrsbeziehungen gegeben. Die in dieser Verkehrsmatrix definierten Verkehrsflüsse dienen als Ausgangspunkt für die drei Dimensionierungsmethoden – nicht robuste, robuste und BGP+-Dimensionierung. Dabei wird angenommen, daß die Verkehrsflüsse ausschließlich Interdomain-QoS-Verkehr repräsentieren, den das NGN von anderen NGN empfängt und/oder an benachbarte NGN versendet. Somit wird der Intradomain-QoS-Verkehr des NGN und der BE-Verkehr bei diesem Vergleich nicht berücksichtigt.

Der Vergleich der drei Dimensionierungsmethoden wird anhand der Summe der benötigten Kapazitäten aller Verbindungen des betrachteten NGN durchgeführt. Dabei weisen die Ergebnisse für die Dimensionierung, bei der das BGP+-Routing berücksichtigt wird, eine wesentlich stärkere Abhängigkeit von der Klassifizierung des NGN und der Einteilung der Interdomain-Verbindungen gegenüber der ausgewählten Topologie auf. Die relativen Einsparungen bei der benötigten Gesamtkapazität eines NGN werden im wesentlichen durch die Reduzierung der absoluten Verkehrsmenge, die durch das NGN transportiert wird, erreicht. Damit ist das Verbesserungspotential durch das Einbeziehen des BGP+-Routing im Vergleich zur Dimensionierung anhand des Schlauchmodells relativ unabhängig von der ausgewählten vermaschten Topologie.

5.3.2 Verwendete Verkehrsmatrizen

Alle Verkehrsmatrizen der verschiedenen Dimensionierungsmethoden bauen auf der in [Zus06] gegebenen Verkehrsmatrix auf. Bei der nicht robusten Dimensionierung wird das Netz für die unveränderte Verkehrsmatrix, bei der jedem Verkehrsfluß eine Quelle und eine Senke fest zugeord-

net ist, ausgelegt. Wegen der 14 fehlenden Verkehrsbeziehungen stellen die 121 symmetrischen Verkehrsflüsse keine vollvermaschte Verkehrsmatrix dar.

Das Schlauchmodell wird bei der robusten Dimensionierung angewendet. Dabei werden alle abgehenden Verkehrsbeziehungen eines Quellknotens zusammengezählt und als ein Schlauch behandelt. Daraus werden dann die Verkehrsbeziehungen des Quellknotens ermittelt, indem zu jedem anderen Knoten ein Verkehrsfluß definiert wird, dessen Menge gleich dem Schlauch ist. Somit wird der eingehende Verkehr jedes Knotens an alle anderen versendet, und es entsteht eine vollvermaschte Verkehrsmatrix. Die empfangene Verkehrsmenge eines Knotens berechnet sich aus der Summe der Schläuche aller anderer Knoten.

Bei der BGP+-Dimensionierung werden jeweils die verallgemeinerten Policies eines Stub-NGN, eines Tier2-NGN und eines Transit-NGN auf die Verkehrsmatrix der robusten Dimensionierung angewendet. Von dieser Verkehrsmatrix wird der zu einem Schlauch zusammengefaßte Eingangsverkehr jedes Knotens weiterverwendet. In Abhängigkeit des betrachteten NGN-Typs wird dieser Verkehr den zur Auswahl stehenden Ausgangs-Router zugewiesen. Damit ergibt sich für jeden NGN-Typ eine unterschiedliche nicht vollvermaschte Verkehrsmatrix.

Für die BGP+-Dimensionierung eines Transit-NGN wurden die Knoten des Beispielnetzes, die alle Grenzknoten zu den benachbarten NGN darstellen, in zwei Kategorien eingeteilt. Als Knoten, die mit einer Peer-to-Peer-Verbindung mit den anderen NGN (Peer-to-Peer-NGN) verbunden sind, wurden die zwei Knoten mit dem größten ausgehenden Verkehr ausgewählt. Die verbleibenden fünfzehn Knoten repräsentieren die Übergänge zu den Kunden-NGN. Die Schläuche des von den Kunden-NGN empfangenen Verkehrs werden ohne Änderung von der robusten Dimensionierung übernommen. Bei dem eingehenden Verkehr über die zwei Peer-to-Peer-Verbindungen muß beachtet werden, daß dieser Verkehr nicht zu dem Router der anderen Peer-to-Peer-NGN weitergeleitet wird und somit der jeweilige Schlauch um einen Ausgangs-Router verringert wird.

Wie schon in Abschnitt 5.2.3 erläutert wurde, muß bei einem Tier2-NGN neben dem Transit-Verkehr auch der von dem Tier2-NGN erzeugte bzw. terminierte Interdomain-QoS-Verkehr betrachtet werden. In diesem Beispiel werden alle siebzehn Knoten als Grenz-Router des NGN eingeordnet und somit existieren keine Router, an denen Interdomain-QoS generiert oder terminiert wird. Aber der Interdomain-QoS-Verkehr, der mit den Kunden-NGN ausgetauscht wird, kann genauso als an dem entsprechenden Grenz-Router generierter bzw. terminierter Interdomain-QoS-Verkehr verstanden werden. Diese Gleichsetzung ändert nichts an den Ergebnissen, da beide Arten von Interdomain-QoS-Verkehr von BGP+ absolut gleich behandelt werden.

Aufgrund dessen wurden bei der Abbildung der Charakteristika eines Tier2-NGN auf das Beispielnetz die Knoten in drei Kategorien unterteilt. Als Knoten mit den Verbindungen mit den Betreiber-NGN wurden die zwei Knoten mit dem größten ausgehenden Verkehr ausgewählt. Die drei Knoten mit den nähst kleineren ausgehenden Verkehrsmengen stellen die Grenz-Router mit den Peer-to-Peer-Verbindungen dar und die restlichen Knoten besitzen die Verbindungen zu den Kunden-NGN.

Der von den Peer-to-Peer-NGN empfangene Interdomain-QoS-Verkehr wird ausschließlich zu den Kunden-NGN übertragen, und dasselbe trifft auf den eingehenden Verkehr von den Betreiber-NGN zu. Damit werden die Schläuche, die an diesen fünf Grenz-Routern beginnen, um vier

Ausgangs-Router im Vergleich zur robusten Dimensionierung reduziert. Wie bei den Transit-NGN ändert sich an den Schläuchen für den eingehenden Verkehr von den Kunden-NGN nichts.

Die charakteristische Eigenschaft eines Stub-NGN sind die zwei Interdomain-Verbindungen. Für das Beispielnetz mit siebzehn Knoten bedeutet das, daß zwei der siebzehn Knoten Ein- bzw. Ausgangs-Router des NGN sind und die restlichen Knoten Quellen und Senken von Interdomain-Verkehr darstellen. Somit kann der Interdomain-Verkehr nur über die beiden Interdomain-Verbindungen mit den anderen NGN ausgetauscht werden. Angewendet auf die Verkehrsmatrix des Stub-NGN bedeutet dies, daß die Verkehrsmenge eines Schlauchs jeweils an die beiden Grenz-Router des Stub-NGN übertragen wird. In der Gegenrichtung wird die Verkehrsmenge der Schläuche der beiden Grenz-Router an alle anderen Knoten weitergeleitet, wobei der jeweils andere Grenz-Router davon ausgenommen ist. Als Grenz-Router des Beispielnetzes wurden dieselben zwei Knoten wie bei der Dimensionierung für ein Transit-NGN bzw. Tier2-NGN ausgewählt, die die zwei größten Verkehrsmengen an die restlichen Knoten versenden.

5.3.3 Eingesetztes Routing-Verfahren

Nachdem die Verkehrsmodellierung inklusive der Bestimmung der Verkehrsmatrizen für die verschiedenen Dimensionierungsmethoden im letzten Abschnitt erläutert wurde, müssen noch weitere Schritte durchgeführt werden, bevor man die benötigte Kapazität der Verbindungen des Beispielnetzes ermittelt kann. Wie aus Abbildung 2.7 ersichtlich ist, müssen dazu noch die Netztopologie geplant, das Routing festgelegt und die Gewährleistung der Fehlerrobustheit sowie der Dienstqualität sichergestellt werden.

Für das hier untersuchte Beispiel ist die Topologie aufgrund des existierenden Netzes schon vorgegeben. Welches Routing-Verfahren zum Einsatz kommt, wird im folgenden beschrieben, und die Mechanismen für die Fehlerrobustheit stehen im Abschnitt 5.3.6 im Vordergrund. Teilweise ist die Bereitstellung der geforderten Dienstqualität schon in die Bestimmung der Verkehrsmatrizen mit eingeflossen, aber weitergehende Aspekte, wie sie in Abschnitt 2.2 überblicksweise diskutierten wurden, werden aus Komplexitätsgründen nicht berücksichtigt.

Aufgrund der sehr großen Verbreitung in heutigen IP-basierten Netzen wurde als Intradomain-Routing-Protokoll OSPF, das in Abschnitt 2.1.1 kurz vorgestellt wurde, ausgewählt. Dieses Link-State-Protokoll berechnet mittels des Dijkstra-Algorithmus den kürzesten Pfad zu einem Ziel und benutzt dazu Verbindungsgewichte, die für jede Verbindung eines Netzes konfiguriert werden können.

Die einfachste Konfigurationsmöglichkeit für die Verbindungsgewichte ist, diese auf einen einheitlichen Wert einzustellen. Somit werden die Pfade, die die wenigsten Verbindungen durchlaufen, bevorzugt. Allerdings lassen sich wesentlich geringere Kapazitätsanforderungen realisieren, wenn man optimierte Verbindungsgewichte verwendet. Wie in [FT00] gezeigt wurde, stellt dies im allgemeinen ein NP-hartes Problem dar und ist damit in binomialer Zeit nicht lösbar. Es wird jedoch in [Rie04] ein Ansatz vorgestellt, der es ermöglicht, für sehr kleine Netze eine optimale Lösung zu finden.

Da das Beispielnetz deutlich mehr Knoten und Verbindungen besitzt als die in [Rie04] ermittelte Obergrenze, wurde eine Heuristik für die Ermittlung der optimalen Verbindungsgewichte gewählt. Der Nachteil an diesem Vorgehen ist, daß eine Heuristik eine minimale Lösung findet, die nicht unbedingt ein globales Minimum darstellen muß. Trotzdem ist der Einsatz einer Heuristik sinnvoll, da dadurch der gesamte Kapazitätsbedarf aller Verbindungen im Vergleich zu der Konfiguration mit einheitlichen Gewichten der Verbindungen wesentlich reduziert werden kann. Diese Aussage wird belegt durch die Arbeiten in [FT00].

5.3.3.1 Verwendung eines genetischen Algorithmus als Heuristik

Die für dieses Beispiel verwendete Heuristik nutzt einen genetischen Algorithmus [Gol89]. Umgesetzt wurde die Heuristik für die Optimierung der Verbindungsgewichte mittels der *Heuristic*- [DG05] und *Graph*-Bibliothek [Gru06]. In [DG05] sind die grundlegenden Schritte eines genetischen Algorithmus schon implementiert und können an die eigenen Bedürfnisse angepaßt werden. Welche Schritte ein genetischer Algorithmus prinzipiell während der Suche nach einem Minimum durchläuft, ist in Abbildung 5.7 zu sehen.

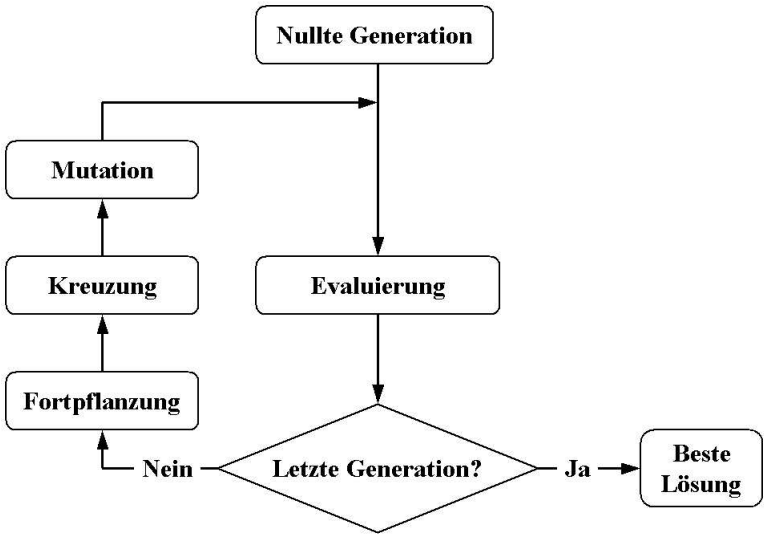


Abbildung 5.7: Flußdiagramm eines genetischen Algorithmus

Der Startpunkt eines genetischen Suchalgorithmus stellt eine Anfangslösung, die in Abbildung 5.7 als die nullte Generation bezeichnet wird, dar, die zufällig gefunden oder vorgegeben wird. Diese nullte Generation besteht aus mindestens zwei verschiedenen Lösungen, welche nach dem Start des genetischen Algorithmus miteinander verglichen werden. Ist dadurch die Abbruchbedingung schon erfüllt, wird die beste Lösung ausgegeben, und der Suchalgorithmus wird beendet. Liefert die Abfrage der Abbruchbedingung eine negative Antwort, werden die in der nullten Generation enthaltenen Lösungen mittels der Methoden Fortpflanzung, Kreuzung und Mutation verändert. Dabei entsteht eine neue Generation von Lösungen, die ebenfalls miteinander verglichen werden. Daran schließt sich wieder die Überprüfung der Abbruchbedingung an. Dieser Ablauf wird

solange wiederholt, bis die Abfrage der Abbruchbedingung positiv ist. Die zu diesem Zeitpunkt beste Lösung wird als Ergebnis des genetischen Algorithmus ausgegeben.

Die Veränderung von einer Generation zur nächsten orientiert sich an der Natur, wo die Lebewesen mit den besseren Eigenschaften eine höhere Überlebenschance besitzen als die mit schlechteren Merkmalen. Bei der Fortpflanzung entsteht aus der aktuellen Generation eine neue, wobei die guten Eigenschaften dieser Generationen von Lösungen mit einer höheren Wahrscheinlichkeit proportional zur Güte dieser Lösungen in die neue Generation übernommen werden. Daran schließt sich die Kreuzung der in der neuen Generation enthaltenen Lösungen an, wobei diese an einer zufälligen Stelle zerteilt und die zweiten Teile zwischen diesen Lösungen ausgetauscht werden. Zum Abschluß der Entwicklung der neuen Generation erfolgt die Mutation. Bei dieser wird versucht, komplett neue Lösungen zu generieren, indem bestimmte Teile einer Lösung zufällig verändert werden. Ziel der Mutation ist es zu vermeiden, daß der genetische Algorithmus nur eine lokal beste Lösung findet und nicht in der Lage ist, weitere Lösungen aufzuspüren.

5.3.3.2 Eingesetzter genetischer Algorithmus

Für die Suche der optimalen Verbindungsgewichte wurde auf eine relativ einfache Implementierung der drei zur Erzeugung neuer Generationen notwendigen Methoden gesetzt, da selbst die besten Methoden nicht verhindern können, daß nur eine lokal beste Lösung gefunden wird, und die komplexeren Algorithmen nur unter besonderen Bedingungen etwas bessere Lösungen finden. Zwei Beispiele für genetische Algorithmen zur Bestimmung der optimalen Verbindungsgewichte, wo die Eigenschaften dieser Probleme bei der Implementierung berücksichtigt werden, werden in [Rie02] und in [RM04] vorgestellt. Dabei wird auch diskutiert, unter welchen Bedingungen diese speziellen Erweiterungen zu besseren Lösungen führen können.

Bei dieser Untersuchung geht es eher darum, die prinzipiellen Kapazitätsgewinne durch die Optimierung der Verbindungsgewichte basierend auf einer Heuristik zu nutzen. Als Parameter für die verwendete Heuristik wurde die Größe einer Generation mit 50 verschiedenen Lösungen ausgewählt. Als Ausgangspunkt der Heuristik muß die nullte Generation vorgegeben werden. Dabei besteht die erste Lösung der nullten Generation aus der Konfiguration, bei der alle Verbindungsgewichte gleich eins sind. Davon ausgehend werden alle weiteren 49 Lösungen der nullten Generation mit zufälligen Verbindungsgewichten ermittelt. Im ersten Schritt des genetischen Algorithmus werden diese Lösungen der nullten Generation miteinander verglichen. Als beste Lösung gilt die, die die kleinste maximale Auslastung aller Verbindungen des Beispielnetzes aufweist.

Da als Abbruchbedingungen 10000 Iterationen bzw. eine maximale Laufzeit von einer Minute definiert wurden, sind beide nach der Festlegung der nullten Generation nicht erfüllt. Deswegen schließt sich an den Vergleich der Lösungen der nullten Generation die Fortpflanzung an. Dabei wird aus der alten Generation eine neue erzeugt. Dazu werden alle Lösungen der alten Generation nach ihrer maximalen Verbindungsauslastung sortiert und die mit den größten Werten gelöscht. Wie viele Lösungen nicht in die neue Generation mit übernommen werden, ist ein weiterer konfigurierbarer Parameter. Bei dieser Untersuchung hat sich ein Wert von eins als aus-

reichend erwiesen. Somit wird bei jeder Fortpflanzung nur die schlechteste Lösung durch eine neue per Zufall erzeugte Lösung ersetzt.

Im Anschluß wird jede Lösung der neu entstandenen Generation mit einer zufällig ausgewählten Lösung derselben Generation gekreuzt. Dabei werden die beiden Lösungen (Elternteil 1 und 2) an einer zufälligen Stelle zerteilt und die abgeteilten Reste untereinander ausgetauscht, wie es in Abbildung 5.8 dargestellt ist.

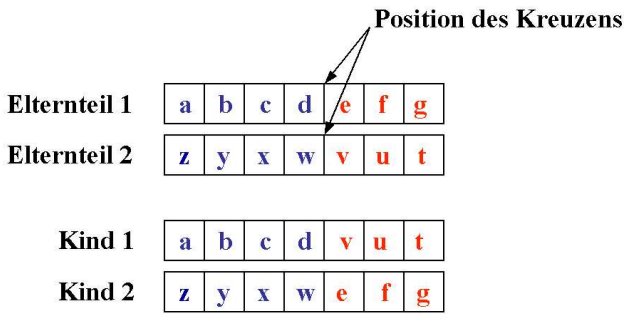


Abbildung 5.8: Kreuzen zweier Lösungen

Somit entstehen beim Kreuzen zweier Elternteile zwei neue Lösungen (Kind 1 und 2). Die so entstandenen Lösungen werden im letzten Schritt noch mutiert. Dazu wird ein Verbindungsgewicht aus der entsprechenden Lösung durch ein zufälliges neues Verbindungsgewicht ersetzt und damit ist die Entwicklung der neuen Generation abgeschlossen. Nach der Bestimmung der neuen besten Lösung werden die Abbruchbedingungen überprüft. Solange eine der beiden Bedingungen nicht erfüllt ist, werden mit den eben beschriebenen Schritten neue Generationen erzeugt und nach einer neuen besten Lösung durchsucht. Da die jeweils besten Lösungen einer Generation immer in die nächste Generation übernommen werden, stellt die beste Lösung der letzten Generation auch die beste bis dahin gefundene Lösung aller Generationen dar.

5.3.4 Ergebnisse der BGP+-Dimensionierung

Benutzt man die im letzten Abschnitt ermittelten Verbindungsgewichte beim Routing mit OSPF, werden die Verkehrsflüsse der Verkehrsmatrix in das NGN gelegt, so daß die maximale Auslastung der Verbindungen minimal ist. Dafür wurde im voraus ein fester Wert für die Kapazität aller Verbindungen festgelegt und anhand der benötigten Kapazität die Auslastung aller Verbindungen ermittelt.

Da bei einer Heuristik nicht sichergestellt ist, daß sie das globale Minimum findet, wurden für jede Verkehrsmatrix die Verbindungsgewichte hundertmal bestimmt. Aufgrund der verschiedenen Zufallswerte der Heuristik wurden in jedem Lauf immer unterschiedliche Ergebnisse gefunden. Die für jede Verkehrsmatrix ermittelte Gesamtkapazität, bei der die geringste maximale Auslastung aller Verbindungen auftritt, ist in Abbildung 5.9 veranschaulicht.

Erwartungsgemäß benötigt die nicht robuste Dimensionierung die geringste Gesamtkapazität. Die Summe der Kapazitäten aller Verbindungen beträgt 3320 Kapazitätseinheiten (KE). Die

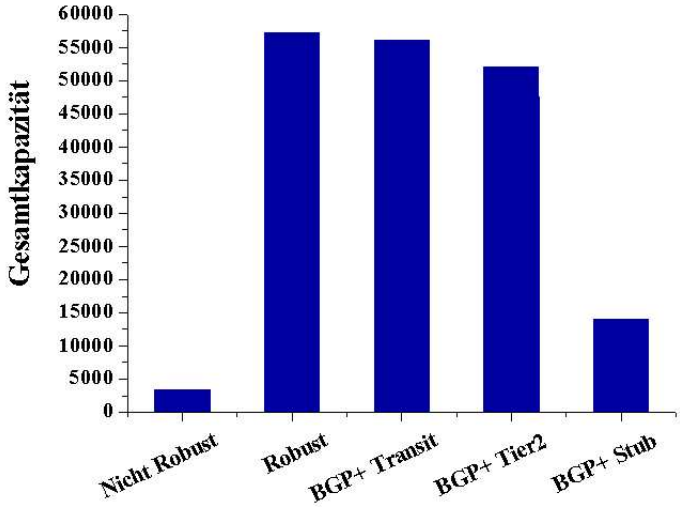


Abbildung 5.9: Kapazitätsbedarf der nicht robusten, der robusten und der drei BGP+-Dimensionierungen

größte Gesamtkapazität ist erforderlich, wenn man die robuste Dimensionierung einsetzt. Im Vergleich zur nicht robusten Dimensionierung wird hier siebzehnmal mehr Kapazität benötigt. Allerdings besteht bei dieser Dimensionierung die Freiheit, den Ausgangs-Router der einzelnen Verkehrsflüsse frei zu wählen.

Da diese Wahlfreiheit aber durch das Interdomain-Routing eingeschränkt wird, ist die starke Überdimensionierung für ein NGN nicht notwendig. Berücksichtigt man das BGP+-Routing eines Transit-NGN, verringert sich die benötigte Gesamtkapazität auf 56128KE. Dies entspricht aber nur einer Reduzierung von 2% im Vergleich zur robusten Dimensionierung. Der Grund für diesen sehr geringen Unterschied liegt darin, daß bei einem Transit-NGN nur zwischen den Ausgangs-Routern, die mit einem anderen Transit-NGN verbunden sind, kein Verkehr ausgetauscht wird. In diesem Beispiel bedeutet dies, daß nur bei zwei Schläuchen ein Ausgangs-Router ausgeschlossen werden konnte, was den geringen Unterschied bei der Gesamtkapazität erklärt.

Bei einem Tier2-NGN können bei den Schläuchen der Ausgangs-Router mit einer Verbindung zu einem Peer-to-Peer- oder Betreiber-NGN die jeweils anderen ausgeschlossen werden und somit liegt die benötigte Gesamtkapazität bei 51998KE, was einer Reduzierung von 9,1% entspricht.

Die geringste Gesamtkapazität mit 14038KE ist bei der BGP+-Dimensionierung für ein Stub-NGN erforderlich, da hier alle Schläuche des ausgehenden Interdomain-QoS-Verkehrs auf zwei Ausgangs-Router beschränkt werden.

Wie diese Ergebnisse der BGP+-Dimensionierung für die drei Arten von NGN zeigen, kann man mit dem Einbeziehen der allgemeinen Routing-Regeln von BGP+ den Kapazitätsbedarf gegenüber der robusten Dimensionierung reduzieren. Allerdings ist nur mit der Dimensionierung für ein Stub-NGN eine wesentliche Einsparung von 75% möglich. Deswegen werden im folgenden Möglichkeiten für die weitere Reduzierung der benötigten Kapazität für die BGP+-Dimensionierung von Transit- und Tier2-NGN untersucht.

Unabhängig davon ist eine weitere Verringerung der Kapazitätsanforderungen möglich, wenn man anstatt der verallgemeinerten Policies die individuellen Routing-Präferenzen der Netzbetreiber in Betracht zieht. Dies kann aber in dieser Arbeit nicht untersucht werden, weil die Policies der Netzbetreiber nicht öffentlich verfügbar sind.

5.3.5 Eingeschränkt-robuste BGP+-Dimensionierung

Die Einsparung des Kapazitätsbedarfs eines NGN bei der im letzten Abschnitt betrachteten BGP+-Dimensionierung beruht auf der Eingrenzung aller möglichen Ausgangs-Router basierend auf den verallgemeinerten Policies. Damit ist gewährleistet, daß selbst in der Ausnahmesituation, in der aller eingehender und erzeugter Interdomain-QoS-Verkehr über einen einzigen Ausgangs-Router an das nächste NGN weitergeleitet wird, genügend Kapazität auf allen Verbindungen zur Verfügung steht. Damit bietet die BGP+-Dimensionierung dieselbe Robustheit wie die robuste Dimensionierung.

Da bei den Transit- bzw. Tier2-NGN aufgrund der geringen Einschränkung der Ausgangs-Router relative kleine Reduzierungen des Kapazitätsbedarfs ermittelt wurden, wird im folgenden die BGP+-Dimensionierung um einige Annahmen erweitert. Diese neuen Annahmen führen dazu, daß die Ausnahmesituationen, in denen aller empfangener Interdomain-QoS-Verkehr sich auf einen Ausgangs-Router konzentriert, nicht mehr berücksichtigt werden. Damit wird mit der als eingeschränkt-robuste BGP+-Dimensionierung bezeichnete Dimensionierung eine Verringerung der Gesamtkapazität auf Kosten der Robustheit erreicht.

Der eingehende Interdomain-QoS-Verkehr eines Eingangs-Routers besteht aus vielen unterschiedlichen Verkehrsflüssen. Im Normalfall werden diese anhand ihrer Zielpräfixe auf die verschiedenen Ausgangs-Router des NGN aufgeteilt. Bei der BGP+-Dimensionierung im letzten Abschnitt wird aber das NGN so ausgelegt, daß es möglich ist, daß der gesamte eingehende Interdomain-QoS-Verkehr eines NGN ein Präfix zum Ziel hat und der gesamte Verkehr über einen Ausgangs-Router weitergeleitet wird. Überträgt man dieses Szenario auf ein Transit-NGN, daß die gesamte Welt umspannen kann, dann bedeutet dies, daß der Großteil des weltweit erzeugten Interdomain-Verkehrs einem einzigen Dienst, der unter einem Präfix erreichbar ist, zuzuordnen ist. Selbst wenn ein Dienst solch ein weltweites Interesse hervorruft, kann der Dienst nicht über ein Präfix erreichbar sein, da solch eine Leistungsfähigkeit nur sinnvoll verteilt angeboten werden kann. Sobald der Interdomain-Verkehr mehrere Präfixe als Ziel hat, kann anhand der verschiedenen Präfixe dieser Verkehr auf unterschiedliche Pfade aufgeteilt werden. Mit diesem Vorgehen wird die extreme Konzentration des Verkehrs auf einen Ausgangs-Router verhindert. Umso stärker der Verkehr mit unterschiedlichen Zielpräfixen auf mehrere Ausgangs-Router verteilt werden kann, umso geringer fällt die Konzentration des Verkehrs und umso unwahrscheinlicher wird die oben beschriebene Extremsituation.

Deswegen ist es angemessen, die Transit- und Tier2-NGN nicht für diese Ausnahmesituationen zu dimensionieren, da diese sehr viel Kapazität benötigen. Um trotzdem genügend Kapazität für die Variabilität der Ausgangs-Router eines NGN bereitzustellen, muß man die Aufteilung des eingehenden Interdomain-QoS-Verkehrs auf die Ausgangs-Router ausreichend robust auslegen.

Für die Bestimmung dieser Aufteilung des eingehenden Interdomain-QoS-Verkehrs können die mit der in Abschnitt 5.2.2 vorgestellten Meßmethode ermittelten Werte weiterverwendet werden. Dafür müssen die in [FGL+01] als transit flow und outbound flow bezeichneten Verkehrsflüsse ausgewertet und die sogenannten *peering links* in die drei Verbindungstypen – Kunden-, Betreiber- und Peer-to-Peer-Verbindung – unterteilt werden. Basierend auf diesen Werten läßt sich ermitteln, wie der eingehende Interdomain-QoS-Verkehr der Kunden-NGN auf die Verbindungen zu den anderen Kunden-NGN, Betreiber-NGN und Peer-to-Peer-NGN aufgeteilt wird. Dasselbe kann auch für den eingehenden Interdomain-QoS-Verkehr der Betreiber-NGN und Peer-to-Peer-NGN bestimmt werden. Mit den Maximalwerten dieser erzielten Aufteilungsverhältnisse kann man die Verkehrsmenge, die auf die verschiedenen Interdomain-Verbindungsarten verteilt wird, einschränken und somit eine Reduzierung des Kapazitätsbedarfs erreichen.

Um zu ermitteln, wie sich der in dem NGN terminierte Interdomain-QoS-Verkehr auf die verschiedenen Router aufteilt, kann derselbe Ansatz eingesetzt werden. Jedoch müssen dafür die als inbound flow bezeichneten Verkehrsflüsse als Basis herangezogen werden.

Welche Kapazitätsersparnis diese eingeschränkt-robuste BGP+-Dimensionierung für ein Transit- bzw. Tier2-NGN bringt, wird im folgenden beispielhaft an zwei verschiedenen Aufteilungsverhältnissen näher untersucht.

5.3.5.1 Eingeschränkt-robuste BGP+-Dimensionierung für ein Transit-NGN

Für diese Untersuchungen wird wieder das Beispiel aus Abschnitt 5.3.4 mit den an die neuen Gegebenheiten angepaßten Verkehrsmatrizen benutzt.

Bei einem Transit-NGN ist auf alle Fälle eine Aufteilung des eingehenden Interdomain-QoS-Verkehrs auf mehrere Ausgangs-Router anzunehmen, da ein Transit-NGN sehr viele verschiedene Verkehrsflüsse weiterleitet, viele Interdomain-Verbindungen besitzt und somit auch sehr viele Präfixe über verschiedene Pfade erreichen kann.

Im ersten Beispiel wird dasselbe Verteilungsverhältnis für alle drei Interdomain-Verbindungstypen angenommen und davon ausgegangen, daß von dem an einem Eingangs-Router eintreffenden Interdomain-QoS-Verkehr nur maximal 75% zu einem einzigen Ausgangs-Router weitergeleitet werden. Die restlichen 25% müssen demnach zu mindestens einem anderen Ausgangs-Router übertragen werden.

Das zweite Beispiel geht davon aus, daß maximal 50% des von den Kunden-NGN empfangenen Interdomain-QoS-Verkehrs zu einem Ausgangs-Router weitergeleitet werden und die andere Hälfte auf andere Ausgangs-Router verteilt wird. Die durch die Annahmen dieser beiden Beispiele erzielten Kapazitätsanforderungen sind in Abbildung 5.10 dargestellt.

Für den besseren Vergleich wurden die Werte für die robuste Dimensionierung ("Robust"), die robuste BGP+-Dimensionierung für ein Transit-NGN ("BGP+ Transit-robust") und für ein Stub-NGN ("BGP+ Stub") aus Abschnitt 5.3.4 mit in Abbildung 5.10 aufgenommen. Die Ergebnisse der beiden Beispiele der eingeschränkt-robusten BGP+-Dimensionierung für ein Transit-NGN werden als "BGP+ Transit-75" bzw. "BGP+ Transit-50" bezeichnet.

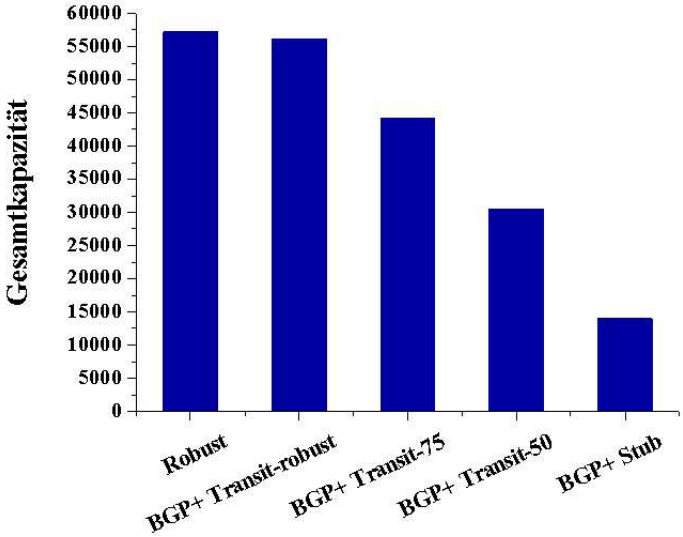


Abbildung 5.10: Benötigte Gesamtkapazität für ein Transit-NGN

Wie in Abbildung 5.10 zu erkennen ist, führt die eingeschränkt-robuste BGP+-Dimensionierung zu einer deutlichen Kapazitätseinsparung. Werden maximal 75% des eingehenden Interdomain-QoS-Verkehrs eines Eingangs-Routers zu einem einzigen Ausgangs-Router übertragen, kann die Gesamtkapazität um 21,3% im Vergleich zu der robusten BGP+-Dimensionierung für ein Transit-NGN verringert werden. Senkt man die Obergrenze für den von den Kunden-NGN empfangenen Interdomain-QoS-Verkehr auf 50%, dann reduziert sich der Kapazitätsbedarf sogar um 45,7%.

5.3.5.2 Eingeschränkt-robuste BGP+-Dimensionierung für ein Tier2-NGN

Bei der Berücksichtigung des Interdomain-Routing eines Tier2-NGN gilt auch, daß genügend Pfade genutzt werden können, um zu verhindern, daß der gesamte eingehende Interdomain-QoS-Verkehr über einen einzigen Ausgangs-Router weitergeleitet wird. Allerdings muß bei einem Tier2-NGN in Betracht gezogen werden, daß alle drei Verbindungstypen existieren und die Anzahl der Nachbar-NGN in einem großen Bereich liegen kann. Deswegen ist für kleine Tier2-NGN zu empfehlen, sie wie Stub-NGN mit der robusten BGP+-Dimensionierung auszulegen.

Für große Tier2-NGN kann die eingeschränkt-robuste BGP+-Dimensionierung angewendet werden, da man hierbei davon ausgehen kann, daß der empfangene Interdomain-QoS-Verkehr eines Eingangs-Routers auf mehrere Ausgangs-Router verteilt werden kann. Im ersten Beispiel wird dieselbe Annahme getroffen wie für das erste Beispiel der eingeschränkt-robusten BGP+-Dimensionierung für ein Transit-NGN. Dabei werden maximal 75% des eingehenden Interdomain-QoS-Verkehrs an einen Ausgangs-Router weitergeleitet (markiert als “BGP+ Tier2-75” in Abbildung 5.11).

Das zweite Beispiel berücksichtigt die stärkere Rolle der Betreiber-NGN, die eine wesentlich größere Erreichbarkeit an Präfixen als die Peer-to-Peer-NGN oder sogar andere Kunden-NGN

bieten können. Aufgrund dessen wird ein kleinerer Teil des von den Kunden-NGN empfangenen Interdomain-QoS-Verkehrs auf die Verbindungen zu den Peer-to-Peer-NGN und anderen Kunden-NGN aufgeteilt. So wird die Verkehrsmatrix des ersten Beispiels übernommen, und der von den Kunden-NGN empfangene Interdomain-QoS-Verkehr wird zu maximal 35% zu den Peer-to-Peer-NGN bzw. zu maximal 25% zu anderen Kunden-NGN verteilt. Dieses Beispiel wird in Abbildung 5.11 als "BGP+ Tier2-35" bezeichnet.

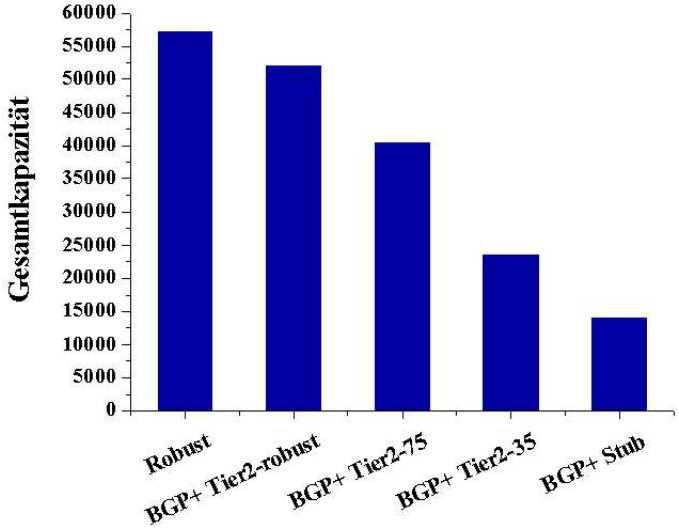


Abbildung 5.11: Benötigte Gesamtkapazität für ein Tier2-NGN

Genau wie in Abbildung 5.10 sind in Abbildung 5.11 auch zu Vergleichszwecken die Ergebnisse der robusten Dimensionierung und der robusten BGP+-Dimensionierung für ein Tier2-NGN und für ein Stub-NGN dargestellt. Auch hier ist eine Reduzierung der Gesamtkapazität durch den Einsatz der eingeschränkt-robusten BGP+-Dimensionierung festzustellen. Das erste Beispiel benötigt 22,2% weniger Kapazität, und der geringere Anteil des von den Kunden-NGN an die Peer-to-Peer-NGN und andere Kunden-NGN weitergeleitete Verkehr führt zu einer Verringerung des Kapazitätsbedarfs um 54,9% im Vergleich zu der robusten BGP+-Dimensionierung für ein Tier2-NGN.

5.3.6 Anwendung der Dimensionierungsregeln des KGID

In den letzten beiden Abschnitten wurde ein NGN für Interdomain-QoS-Verkehr dimensioniert und dabei angenommen, daß die Menge des Interdomain-QoS-Verkehr sich nicht ändert. Deswegen werden nachfolgend die Auswirkungen von Interdomain-Routing-Änderungen auf die Dimensionierung des NGN näher untersucht.

Die Verkehrsverschiebungen aufgrund einer Interdomain-Routing-Änderung wurden im Kapitel 4 detailliert analysiert und daraus Dimensionierungsregeln für die Interdomain-Verbindungen

zwischen den NGN abgeleitet. Diese Vorschriften werden nun auf die Dimensionierung eines NGN angewendet, wobei wieder zwischen den verschiedenen Typen an NGN und deren Interdomain-Verbindungen unterschieden wird.

Um die Interdomain-Verbindungen eines Transit-NGN ausreichend für die Verkehrsverschiebungen, hervorgerufen durch eine BGP+-Routing-Änderung, zu dimensionieren, wird in Abschnitt 4.4.2.5 eine Überdimensionierung von 5% für die Peer-to-Peer-Verbindungen und von 10% bzw. 30% für die restlichen Verbindungen definiert. Im Fall der Tier2-NGN werden für die Peer-to-Peer-Verbindungen ebenfalls 5% und für die verbleibenden Verbindungen 30% mehr Kapazität gefordert.

Diese Überdimensionierung wurde auf die Verkehrsmatrizen der robusten und eingeschränkt-robusten BGP+-Dimensionierung übertragen, indem die eingehende Verkehrsmenge der einzelnen Grenz-Router um den entsprechenden Anteil erhöht wurde. Nach dem Durchlauf der Netzplanung für das Beispielnetz ergibt sich die vergrößerte Gesamtkapazität, wie sie in Abbildung 5.12 zum Ausdruck kommt.

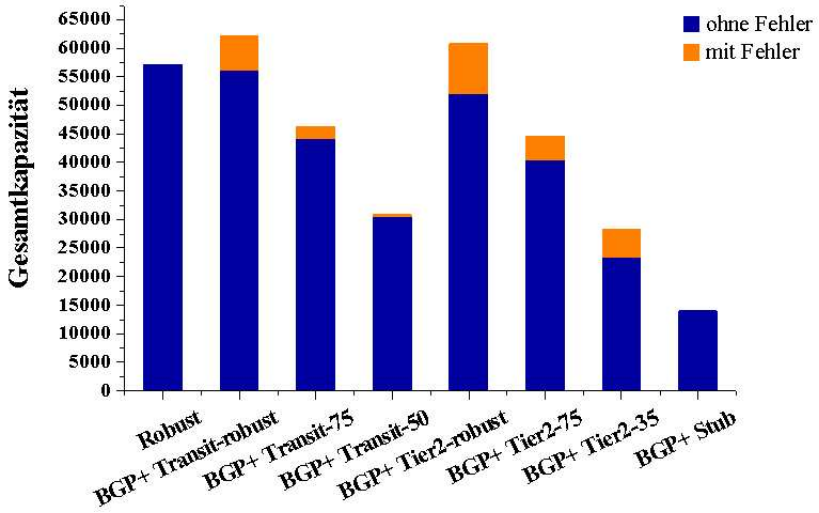


Abbildung 5.12: BGP+-Dimensionierung mit Berücksichtigung der Interdomain-Fehler

Zum Vergleich des Kapazitätsbedarfs der einzelnen Dimensionierungen ist in Abbildung 5.12 auch die benötigte Gesamtkapazität der robusten Dimensionierung zu sehen. Wendet man die Dimensionierungsregeln für ein Transit-NGN an, werden bei der robusten BGP+-Dimensionierung 10,1% mehr Gesamtkapazität für das Beispielnetz benötigt und damit mehr als für die robuste Dimensionierung. Für die eingeschränkt-robuste BGP+-Dimensionierung für ein Transit-NGN ergeben sich Erhöhungen von 4,8% bzw. 1,3%, womit diese sehr moderat ausfallen und deutlich unter der benötigten Gesamtkapazität der robusten BGP+-Dimensionierung liegen. Ein Teil der geringeren Vergrößerung des Kapazitätsbedarfs liegt an der kleineren Verkehrsmenge, die das NGN übertragen muß. Möglicherweise kann ein Teil des Unterschieds auch auf den Einsatz der Heuristik zurückgeführt werden, da die Ergebnisse davon abhängen, wie nahe das von der Heuristik gefundene lokale Minimum am jeweiligen globalen Minimum liegt.

Bei der Dimensionierung für ein Tier2-NGN fällt die Erhöhung der Kapazitätsanforderungen etwas größer aus. So wird bei der robusten Dimensionierung für ein Tier2-NGN 16,9% mehr Kapazität benötigt. Diese stärkere Vergrößerung hängt aber mit den Dimensionierungsregeln für das KGID zusammen. Während bei einem Transit-NGN für die Verbindungen zu den Peer-to-Peer-NGN nur 5% mehr Kapazität nötig sind, werden bei einem Tier2-NGN 30% mehr Kapazität benötigt, da die Peer-to-Peer-Verbindungen der Transit-NGN bei den Tier2-NGN die Verbindungen zu den Betreiber-NGN darstellen. Daraus resultiert der größere Kapazitätsbedarf bei den Tier2-NGN.

Dasselbe Verhalten zeigt sich auch bei der eingeschränkt-robusten BGP+-Dimensionierung, wo sich die Kapazitätsanforderungen um 10,2% bzw. 20,8% vergrößern. Die deutlich größere Erhöhung des zweiten Beispiels ist auf den Einfluß der Heuristik zurückzuführen. Betrachtet man die anderen Lösungen der Heuristik, die eine marginal höhere minimale Maximalauslastung aller Verbindungen aufweisen, ist für diesen Fall eher ein erhöhter Kapazitätsbedarf im Bereich zwischen 8% und 10% zu erwarten.

Neben dieser Vergrößerung zeigt Abbildung 5.12 auch, daß bei einem Stub-NGN keine zusätzliche Kapazität notwendig ist. Die Dimensionierungsregel aus Abschnitt 4.4.2.5 besagt, daß die Interdomain-Verbindungen eines Stub-NGN so ausgelegt werden müssen, daß jede Verbindung den kompletten Interdomain-Verkehr transportieren kann, was in der robusten BGP+-Dimensionierung für ein Stub-NGN schon enthalten ist.

5.3.7 Änderung bei der Netzplanung eines NGN

Nachdem gezeigt wurde, daß die Berücksichtigung des BGP+-Routing bei der Netzdimensionierung eine Effizienzsteigerung aufgrund des reduzierten Kapazitätsbedarfs ermöglicht, wird im folgenden noch auf die Änderungen bei der Netzplanung eingegangen, die durch diese neue Netzdimensionierung entstehen. Am eigentlichen Ablauf der Netzplanung, wie er in Abschnitt 2.3 beschrieben wurde, ändert sich nichts. Nur bei der Bestimmung der Verkehrsmatrix gilt es, die neuen Dimensionierungsverfahren anzuwenden.

Ausgangspunkt der BGP+-Dimensionierungsmethoden ist die Unterteilung in BE-Verkehr und QoS-Verkehr sowie in Verkehr, der das betrachtete NGN nicht verläßt, und den Interdomain-Verkehr. Nur für den Interdomain-Verkehr, für den die Dienstqualität garantiert werden muß, bringt die Anwendung des neuen Dimensionierungsverfahren Vorteile. Die Auslegung eines NGN für die restlichen drei Verkehrsarten kann wie gehabt mit der sehr kapazitätseffizienten, nicht robusten Dimensionierung erledigt werden.

Bei den BGP+-Dimensionierungsverfahren wird davon ausgegangen, daß der Netzbetreiber des NGN weiß, an welchem Eingangs-Router die Kunden ihren Interdomain-QoS-Verkehr ins Netz einspeisen werden und wieviel Verkehr dies ist. Anhand dieser bekannten Eingangsverkehre können die Verkehrsmengen aller Schläuche ermittelt werden und darauf die BGP+-Dimensionierungsverfahren angewendet werden. Basierend auf den so erzielten Ergebnissen können die Interdomain-Verbindungen dimensioniert werden, wobei hier die Dimensionierungsregeln

für das KGID berücksichtigt werden müssen (s. Abschnitt 4.4.2.5). Diese Verkehrsmenge repräsentiert den eingehenden Interdomain-QoS-Verkehr der Nachbar-NGN und bildet somit den Ausgangspunkt für dessen Netzdimensionierung.

Dieser Prozeßablauf führt zu einer Abhängigkeit, wo ein Netzbetreiber erst nach der Dimensionierung seiner Kunden-NGN den eingehenden Interdomain-QoS-Verkehr seines NGN kennt. Aufgrund dessen ergibt sich ein quasi hierarchischer Ablauf der Netzplanung, der bei den Stub-NGN beginnt. Da diese nur Verkehr generieren bzw. terminieren, sind sie in der Lage, unabhängig von anderen NGN ihren Interdomain-QoS-Verkehr bestimmen zu können. Diese Verkehrsmengen bilden die Grundlage für die Betreiber-NGN der Stub-NGN, und deren ausgehender Interdomain-QoS-Verkehr ist der Ausgangspunkt für die Netzdimensionierung für deren Betreiber-NGN. Diese Abhängigkeit setzt sich fort bis zur Netzplanung der Transit-NGN.

Daraus darf aber nicht der Schluß gezogen werden, daß die größeren NGN immer mit ihrer Netzplanung auf ihre Kunden-NGN warten müssen. Vielmehr führt diese Abhängigkeit dazu, daß ein Netzbetreiber mit sehr vielen Kunden die Netzdimensionierung seines NGN häufiger ausführen muß.

5.4 Zusammenfassung

Nachdem in Kapitel 4 ein Konzept für den Transport von Interdomain-QoS-Verkehr zwischen verschiedenen NGN erarbeitet wurde, standen in diesem Kapitel die daraus resultierenden Anforderungen an die Netzplanung eines NGN im Vordergrund. Mit der Übertragung von Interdomain-QoS-Verkehr übernimmt ein Netzbetreiber die Verantwortung sicherzustellen, daß der Interdomain-QoS-Verkehr sein Netz passiert, ohne dabei die Dienstqualität zu beeinträchtigen. Diese geänderten Anforderungen im Vergleich zum Datentransport von Best-Effort-Verkehr schlagen sich auch bei der Netzplanung nieder, da diese das Werkzeug darstellt, mit der die neuen Herausforderungen gelöst werden können.

Bei der Untersuchung bestehender Netzplanungsverfahren hat sich herausgestellt, daß die nicht robuste Dimensionierung für den Interdomain-QoS-Verkehr nicht geeignet ist. Die robuste Dimensionierung basierend auf dem Schlauchmodell erfüllt die Anforderungen, aber daraus resultiert eine sehr ineffiziente Dimensionierung des NGN, da hierbei sehr viel Kapazität benötigt wird. Deswegen wurde ein neues Netzdimensionierungsverfahren entworfen, das den Anforderungen des Interdomain-Transports von QoS-Verkehr Rechnung trägt.

Ausgehend von der robusten Dimensionierung wurden die allgemeinen Policies von BGP+ benutzt, um die uneingeschränkte Auswahl des Ausgangs-Routers bei der robusten Dimensionierung einzugrenzen. Da das BGP+-Routing entscheidet, welche Ausgangs-Router für das Erreichen eines Ziel in einem anderen NGN ausgewählt werden, stellt diese Einschränkung keine Reduzierung der Wahlfreiheit dar, sondern die vorhandenen und von einem Netzbetreiber festgelegten Beschränkungen werden bei der Netzdimensionierung berücksichtigt.

Anhand eines Beispiels mit einer gegebenen Topologie wurden die verschiedenen Dimensionierungsverfahren miteinander verglichen. Dabei hat sich gezeigt, daß durch das Einbeziehen des

BGP+-Routing bei der Bestimmung der Verkehrsmatrix eines NGN deutliche Einsparungen bei der benötigten Kapazität aller Verbindungen eines Netzes zu erreichen sind. Bei der Beispielrechnung konnte in Abhängigkeit des Typs des betrachteten NGN der Kapazitätsbedarf in einem Bereich von bis zu 75% im Vergleich zu der robusten Dimensionierung verringert werden. Allerdings war diese deutliche Verbesserung nur für die Stub-NGN zu erreichen.

Deswegen wurde für die Transit- und Tier2-NGN die robuste BGP+-Dimensionierung auf Kosten der Robustheit erweitert, indem die Aufteilung des eingehenden Interdomain-QoS-Verkehrs auf die Ausgangs-Router in Abhängigkeit von deren Interdomain-Verbindungstypen bei der Dimensionierung berücksichtigt wurde. Mit dieser sogenannten eingeschränkt-robusten BGP+-Dimensionierung konnte in den vier Beispielen der Kapazitätsbedarf in einem Bereich von 20% bis 55% reduziert werden.

Wendet man die Dimensionierungsregeln für die KGID an, erhöht sich der Kapazitätsbedarf der Transit- und Tier2-NGN. Bei den Stub-NGN sind die Anforderungen der Interdomain-Verkehrsverschiebungen schon mit der robusten BGP+-Dimensionierung erfüllt. Für die Tier2-NGN und Transit-NGN wurde in der Beispielrechnung ein Mehrbedarf in einem Bereich von 1,3% bis 20,8% mehr Kapazität festgestellt.

Mit diesen neuen Dimensionierungsmethoden kann ein NGN so ausgelegt werden, daß es neben den Verkehrsverschiebungen aufgrund von Interdomain-Routing-Änderungen auch die Variabilität der Ausgangs-Router eines NGN ohne Beeinträchtigung der Dienstqualität unterstützt. Damit wurde in diesem Kapitel aufgezeigt, wie man die Anforderungen des Interdomain-QoS-Verkehrs bei der Netzplanung eines NGN erfüllen kann.

Kapitel 6

Zusammenfassung

Hervorgerufen durch die Konvergenz der Netze ist die Sicherstellung der Dienstqualität in IP-Netzen unumgänglich. Für die Gewährleistung der Dienstqualität innerhalb eines IP-Netzes gibt es die verschiedensten Lösungansätze, die zum Teil schon als Produkte auf dem Markt sind. Die Garantie der Dienstqualität zwischen den IP-Netzen steht im Gegensatz dazu noch am Anfang der Entwicklung. Deswegen wurde in dieser Arbeit untersucht, wie die Interdomain-Dienstqualität kombiniert mit einer hohen Verfügbarkeit gewährleistet werden kann. Dabei war die Zielvorstellung, daß das heutige Internet sich in einem längeren Zeitraum zu einem Netz entwickelt, wo alle autonomen Systeme die Intra- und Interdomain-Dienstqualität garantieren können.

Zu der Sicherstellung der Dienstqualität gehört auch, daß alle dafür notwendigen Mechanismen eine schnelle Fehlerreaktion bieten. Ist dies nicht der Fall, kann die Dienstqualität aufgrund eines Fehlers beeinträchtigt werden. Die sehr lange Dauer der Fehlerreaktion von BGP kann sogar zu einem Abbruch der Datenübertragung von QoS-Verkehr führen. Aus diesem Grund wurde im ersten Teil der Arbeit das BGP-Konvergenzverhalten umfassend analysiert. Während dieser Untersuchung wurde festgestellt, daß nur die Beschleunigung des Verteilens der Routing-Informationen die BGP-Konvergenzzeit ohne Einschränkungen reduzieren kann. Die Erweiterung von BGP um das Enhanced Path Vector Protocol, womit das BGP-Pfadtesten vermieden werden kann, ermöglicht das Deaktivieren des BGP-Zeitglieds MRAI. Mit diesem Vorgehen konnte die BGP-Konvergenzzeit in allen Fehlerszenarien minimiert werden. Anhand einer theoretischen Abschätzung der unteren Grenze der BGP-Konvergenzzeit wurde gezeigt, daß das Minimum der ermittelten BGP-Konvergenzzeit genau dieser unteren Grenze entspricht.

Die Erkenntnisse wurden in die neue Erweiterung von BGP, in das BGP+, integriert und alle notwendigen Änderungen im Vergleich zu BGP spezifiziert. Damit wird sichergestellt, daß BGP+ und BGP nebeneinander in verschiedenen autonomen Systemen eingesetzt werden können. Wegen dieser Eigenschaft ist es möglich, BGP+ im heutigen Internet schrittweise einzuführen. Allerdings erreicht die Konvergenzzeit des Interdomain-Routing erst ihr Minimum, wenn alle autonomen Systeme BGP+ einsetzen.

Basierend auf BGP+ wurde im zweiten Teil dieser Arbeit das Konzept für Garantierte Interdomain-Dienstqualität (KGID) entworfen. Dabei wurde ein Szenario aufgezeigt, daß das all-

mähliche Entstehen von Next Generation Networks (NGN) und deren Verbindung untereinander aufzeigt. Diese Evolution des heutigen Internets ermöglicht es, in und zwischen den NGN neue Mechanismen für die Gewährleistung von Intra- und Interdomain-Dienstqualität einzuführen, ohne die bisherige Best Effort-Datenübertragung zu beeinflussen.

Im Rahmen der Erarbeitung des KGID wurde BGP+ an die Anforderungen des Interdomain-QoS-Verkehrs angepaßt, indem ein weiteres optionales Attribut eingeführt wurde, um zwischen Pfaden zu unterscheiden, die die Dienstqualität zusichern können oder nicht. Eine weitere notwendige Erweiterung stellt die Schnittstelle zum Ressourcenreservierungsprotokoll QoS-NSLP aus der NSIS-Protokollfamilie dar. Diese Schnittstelle ermöglicht die Initiierung der Aktualisierung der Interdomain-Ressourcenreservierungen durch BGP+ zu dem Zeitpunkt, wo sichergestellt ist, daß BGP+ sicher konvergiert ist.

Der Schwerpunkt beim Entwurf des KGID lag auf der Sicherstellung der Interdomain-Dienstqualität auch im Fall einer Interdomain-Topologieänderung. Da aktuelle Informationen über die freien Interdomain-Ressourcen nicht verfügbar sind und das Interdomain-Routing aufgrund der Policies nicht vorhersagbar ist, stellt die einzige Lösung die gezielte Überdimensionierung der Interdomain-Verbindungen dar. Um für diesen Fall Dimensionierungsregeln aufstellen zu können, wurde anhand eines Modells der durch die Interdomain-Topologieänderungen verschobene Interdomain-QoS-Verkehr abgeschätzt. Aus den Ergebnissen resultiert, daß die Interdomain-Verbindungen der Stub-NGN für den gesamten Interdomain-Verkehr ausgelegt werden müssen und bei den Tier2- und Transit-NGN eine Überdimensionierung der Interdomain-Verbindungen in Abhängigkeit des Verbindungstyps von 5%, 10% oder 30% notwendig ist.

Desweiteren wurde eine neue Interdomain-QoS-Klasse mit der höchsten Priorität eingeführt, damit der verschobene Interdomain-QoS-Verkehr, für den keine gültigen Reservierungen existieren, von den Netzzugangskontrollverfahren nicht blockiert und somit die Dienstqualität so gering wie möglich beeinträchtigt wird.

Die Übertragung von Interdomain-QoS-Verkehr führt zu einem Paradigmenwechsel innerhalb eines Netzes. Ein Netzbetreiber garantiert in diesem Fall, daß der Interdomain-QoS-Verkehr das NGN ohne Beeinträchtigung der Dienstqualität passiert. Dies muß bei der Netzplanung eines NGN, das die Interdomain-Dienstqualität sicherstellen will, berücksichtigt werden. Da die bestehenden Verfahren entweder sehr große Kapazitätsanforderungen besitzen oder die Dienstqualität beeinträchtigen, wurde in dieser Arbeit eine neue Methode für die Netzdimensionierung vorgeschlagen. Die neue BGP+-Dimensionierung berücksichtigt das BGP+-Routing bei der Bestimmung der Verkehrsmatrix und ermöglicht so eine Reduzierung der benötigten Gesamtkapazität eines NGN.

Für das untersuchte Beispielnetz wurde eine Verringerung des Kapazitätsbedarfs von 2% bis 75% gegenüber dem bestehenden Ansatz der robusten Dimensionierung erzielt. Für die Fälle, wo die Unterschiede sehr gering ausfallen, wurde die BGP+-Dimensionierung erweitert, so daß die sehr seltenen Fälle mit einem sehr großen Kapazitätsbedarf nicht mehr unterstützt werden. Dabei konnten die Kapazitätsanforderungen um 20% bis 55% reduziert werden.

Anhand des Beispiels konnte gezeigt werden, daß durch die Beachtung des BGP+-Routing bei der Bestimmung der Verkehrsmatrix, der Kapazitätsbedarf eines NGN deutlich verringert werden

kann. Da die BGP+-Dimensionierung die Menge des durch das NGN geleiteten Interdomain-QoS-Verkehrs verkleinert, sind ähnliche Einsparungspotentiale auch für andere Netztopologien zu erwarten.

Mit diesem neuen Konzept wurde ein Weg aufgezeigt, wie man das heutige Internet weiterentwickeln kann, um in und zwischen allen autonomen Systemen die Dienstqualität sowie die Fehlerrobustheit gewährleisten zu können. Dabei wurde das heutige Interdomain-Routing entscheidend verbessert, ein Konzept für die Einführung der neuen Protokolle und Mechanismen entworfen und für einen Netzbetreiber Dimensionierungsregeln und ein Verfahren entworfen, um dessen Netz für die Garantie der Interdomain-Dienstqualität auszulegen.

Ein Aspekt, der in zukünftigen Untersuchungen betrachtet werden sollte, sind die Mechanismen, die notwendig sind, um die Sicherheit der Interdomain-Datenübertragung von QoS-Verkehr zu gewährleisten. Die verschiedenen, bestehenden Möglichkeiten, wie die BGP-Kommunikationen zwischen den autonomen Systemen abgesichert werden kann, werden in [ZSN05] diskutiert und bewertet. Allerdings stellt dies nur den ersten Schritt dar, darüber hinaus muß sichergestellt werden, daß die zugesicherte Dienstqualität eines Verkehrsflusses nicht von einem anderen mißbraucht wird. Ein Konzept, wie man dieses Problem lösen kann, wird in [RRT06] beschrieben. Dieser Ansatz könnte auch als Basis für eine Lösung dienen, um den Mißbrauch der neu eingeführten QoS-Klasse zu verhindern. In diesem Fall kann ein NGN ohne vorherige Ankündigung seinen Nachbarn QoS-Verkehr schicken, der für einen begrenzten Zeitraum der höchsten Interdomain-QoS-Klasse zugeordnet wird.

Abbildungsverzeichnis

1.1 Intra- und Interdomain-Kommunikation 2

2.1 BGP-Zustandsdiagramm nach [Hal97] 8
2.2 Aufbau einer BGP-UPDATE-Nachricht nach [Hal97] 9
2.3 Bestimmung der besten Pfade in einem BGP-Router nach [Hal97] 12
2.4 Netz mit drei BGP-Routern ohne (a) bzw. mit Routen-Reflektor (b) 15
2.5 Netz mit siebzehn BGP-Routern 15
2.6 Netz mit 48 BGP-Routern 16
2.7 Netzplanungszyklus 21
2.8 Detaillierter Netzplanungszyklus 24

3.1 Wachstum des Internets aus Sicht des Betreibers Telstra [Tel05] 32
3.2 Meßwerte der BGP-Konvergenzzeit nach [LAB+01] 36
3.3 BGP-Pfadtesten im Fehlerfall 37
3.4 Zeitverhalten des RFD 40
3.5 Fehlerreaktion mittels FaSRo 42
3.6 Ausfall einer Intradomain- (rot) und Interdomain- (braun) Verbindung 43
3.7 Änderungen eines BGP-Ersatzpfads nach einer Topologieänderung 45
3.8 Durchlaufzeit einer BGP-UPDATE-Nachricht durch ein autonomes System . . . 48
3.9 Modell der Durchlaufzeit eines autonomen Systems mit Routen-Reflektor 50
3.10 Modellierung der Fehlerbilder in SSFnet (a) t_{down}; (b) t_{long} 55
3.11 Konvergenzzeit (a) und Nachrichtenanzahl (b), wenn ein Präfix nicht mehr erreichbar ist 57
3.12 Konvergenzzeit (a) und Nachrichtenanzahl (b), wenn ein Präfix wieder erreichbar ist 59

3.13 Konvergenzzeit (a) und Nachrichtenanzahl (b) beim Umschalten auf einen Ersatzpfad . . . 61
3.14 Vergleich des Konvergenzverhaltens von t_{short} und t_{up} anhand der Zeit (a) und der Nachrichtenanzahl (b) . . . 62
3.15 Einfluß verschiedener Verarbeitungszeiten anhand der Zeit (a) und der Nachrichtenanzahl (b) . . . 64
3.16 Einfluß der internen Topologie eines autonomen Systems anhand der Konvergenzzeit t_{down} (a) und der Nachrichtenanzahl (b) . . . 66
3.17 Einfluß der internen Topologie eines autonomen Systems anhand der Konvergenzzeit t_{up} (a) und der Nachrichtenanzahl (b) . . . 67
3.18 Konvergenzzeit von EPIC, wenn ein Präfix nicht mehr (a) und wieder (b) erreichbar ist . . . 71
3.19 Konvergenzzeit (a) und Nachrichtenanzahl (b) von EPIC und BGP beim Umschalten auf einen Ersatzpfad . . . 72
3.20 Konvergenzzeit (a) und Nachrichtenanzahl (b) von EPIC und BGP beim Umschalten auf einen besseren Pfad . . . 73
3.21 Format der BGP+-UPDATE-Nachricht . . . 75
3.22 Format der BGP-OPEN-Nachricht nach [Hal97] . . . 75

4.1 QoS-Inseln . . . 78
4.2 Auswahl der QoS-Klassen eines Pfads beim MESCAL-Konzept [MBA+04] . . . 81
4.3 Ressourcenreservierung mit BGRP [SGV03] . . . 84
4.4 NSIS Protokollfamilie aus [FSB+05] . . . 86
4.5 NSIS Ende-zu-Ende Kommunikation . . . 86
4.6 Interdomain-Verkehrsverschiebungen . . . 96
4.7 Verteilung der Anzahl der Nachbar-NGN . . . 97
4.8 Bezeichnung der Verbindungen eines NGN . . . 99
4.9 Verkehrsverschiebungen bei NGN82 mit Stub-NGN . . . 102
4.10 Verkehrsverschiebungen bei NGN82 ohne Stub-NGN . . . 103
4.11 Verkehrsverschiebungen bei NGN213 mit (a) bzw. ohne (b) Stub-NGN . . . 103
4.12 Verkehrsverschiebungen bei NGN10 mit Stub-NGN . . . 105
4.13 Verkehrsverschiebungen bei NGN10 ohne Stub-NGN . . . 106
4.14 Verkehrsverschiebungen bei NGN41 mit Stub-NGN . . . 107
4.15 Verkehrsverschiebungen bei NGN41 ohne Stub-NGN . . . 108

4.16 Verkehrsverschiebungen von NGN3 109

5.1 Das Schlauchmodell 117
5.2 Kapazitätsbedarf der statistischen Dimensionierung [Sha05b] 119
5.3 Charakterisierung der Verkehrsflüsse [FGL+01] 121
5.4 Verkehrsflüsse durch ein Transit-NGN 122
5.5 Verkehrsflüsse durch ein Tier2-NGN 123
5.6 Verkehrsflüsse durch ein Stub-NGN 123
5.7 Flußdiagramm eines genetischen Algorithmus 127
5.8 Kreuzen zweier Lösungen 129
5.9 Kapazitätsbedarf der nicht robusten, der robusten und der drei BGP+-Dimensionierungen 130
5.10 Benötigte Gesamtkapazität für ein Transit-NGN 133
5.11 Benötigte Gesamtkapazität für ein Tier2-NGN 134
5.12 BGP+-Dimensionierung mit Berücksichtigung der Interdomain-Fehler 135

Tabellenverzeichnis

2.1 Routing-Einschränkungen in Abhängigkeit der Geschäftsbeziehungen 13

3.1 Standardeinstellungen der RFD 41

3.2 Konvergenzzeit mit beiden MRAI, mit E-BGP-MRAI, ohne MRAI und mit einem Routen-Reflektor 52

4.1 Verkehrsverschiebungen der Stub-NGN 100

4.2 Verkehrsverschiebungen der Tier2-NGN mit bis zu 10 Nachbarn 100

4.3 Verkehrsverschiebungen der kleinen Tier2-NGN ohne Berücksichtigung der Stub-NGN 101

4.4 Verkehrsverschiebungen der Tier2 NGN mit mehr als 10 Nachbarn 104

4.5 Verkehrsverschiebungen der großen Tier2-NGN ohne Berücksichtigung der Stub-NGN 105

4.6 Verkehrsverschiebungen der Transit-NGN 108

4.7 Verkehrsverschiebungen der Transit-NGN ohne Berücksichtigung der Stub-NGN 109

Literaturverzeichnis

Literaturangaben, deren Verweis im Fettdruck erscheint, sind Vorveröffentlichungen des Verfassers, die als Haupt- oder Mitautor erstellt wurden.

[AB00] ALBERT, R. und A. BARABASI: *Topology of Evolving Network: Local Events and Universality*. Physical Review Letters, 85(24):5234–5237, Dezember 2000.

[ACK03] AGARWAL, S., C. CHUAH und R. KATZ: *OPCA: Robust Interdomain Policy Routing and Traffic Control*. In: *Proceedings of the IEEE Openarch*, April 2003.

[ACL00] AIELLO, W., F. CHUNG und L. LU: *A Randowm Graph Model for Massive Graphs*. In: *Proceedings of the 32rd Annual ACM Symposium on Theory of Computing*, 2000.

[AGLAB94] ALBRIHTSON, R., J. GARCIA-LUNA-ACEVES und J. BOYLE: *EIGRP-A Fast Routing Protocol Based on Distance Vectors*. In: *Proceedings of the Networld/Interop*, May 1994.

[AKZ99a] ALMES, G., S. KALIDINDI und M. ZEKAUSKAS: *A One-way Delay Metric for IPPM*. RFC 2679, IETF, September 1999.

[AKZ99b] ALMES, G., S. KALINDINDI und M. ZEKSUSKAS: *A One-Way Packet Loss Metric for IPPM*. RFC 2680, IETF, September 1999.

[Amm05] AMMAR, M.: *Why we STILL Don't Know How to Simulate Networks*. Keynote at the 13th Annual Meeting of the IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), 2005.

[APS+04] AKELLA, A., J. PANG et al.: *A Comparison of Overlay Routing and Multihoming Route Control*. In: *Proceedings of ACM SigComm04*, August 2004.

[Aut02] AUTENRIETH, A.: *Differentiated Resilience in IP-Based Multilayer Transport Networks*. Dissertation, TU München, Institute of Communication Networks, September 2002.

[AVG+99] ALAETTINOGLU, C., C. VILLAMIZAR et al.: *Routing Policy Specification Language (RPSL)*. RFC 2622, IETF, Juni 1999.

[AZ05] ATLAS, A. und A. ZININ: *Basic specification for IP fast reroute: Loop-free alternates*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-ietf-rtgwg-ipfrr-spec-base-04.txt, 2005.

[BA99] BARABASI, A. und R. ALBERT: *Emergence of Scaling in Random Networks*. Science, 286:509–512, Oktober 1999.

[BAK05] BEN-AMEUR, W. und H. KERIVIN: *Routing of Uncertain Traffic Demands*. Optimization and Engineering, 6:283–313, 2005.

[BBAS03] BREMLER-BARR, A., Y. AFEK und S. SCHWARZ: *Improved BGP Convergence via Ghost Flushing*. In: *Proceedings of IEEE Infocom*, March 2003.

[BBC+98] BLAKE, S., D. BLACK et al.: *An Architecture for Differentiated Service*. RFC 2475, IETF, Dezember 1998.

[BCC06] BATES, T., E. CHEN und R. CHANDRA: *BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)*. RFC 4456, IETF, April 2006.

[BCS94] BRADEN, R., D. CLARK und S. SHENKER: *Integrated Services in the Internet Architecture: an Overview*. RFC 1633, IETF, Juni 1994.

[BCS+00] BRADEN, R., D. CLARK et al.: *Developing a Next-Generation Internet Architecture*. Internal Whitepaper, MIT, Laboratory for Computer Science, Juli 2000.

[BFF05] BONAVENTURE, O., C. FILSFILS und P. FRANCOIS: *Achieving sub-50 milliseconds recovery upon BGP peering link failures*. In: *Proceedings of the International Conference On Emerging Networking Experiments And Technologies*, October 2005.

[BFY+00] BERNET, Y., P. FORD et al.: *A Framework for Integrated Services Operation over DiffServ*. RFC 2998, IETF, November 2000.

[BGM02] BUSH, R., T. GRIFFIN und Z. MAO: *Route Flap Damping: Harmful?* Talk at the NANOG 26 meeting, The North American Network Operators' Group - NANOG, Oktober 2002.

[BGS+01] BERGER, L., D. GAN et al.: *RSVP Refresh Overhead Reduction Extensions*. RFC 2961, IETF, April 2001.

[BILF+01] BAKER, F., C. ITURRALDE et al.: *Aggregation of RSVP for IPv4 and IPv6 Reservations*. RFC 3175, IETF, September 2001.

[BLS+03] BLESS, R., G. LICHTWALD et al.: *Fast Scoped Rerouting for BGP*. In: *Proceedings of the 11th IEEE International Conference on Networks (ICON)*, September 2003.

[Bou05] BOUCADAIR, M.: *QoS-Enhanced Border Gateway Protocol*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-boucadair-qos-bgp-spec-01.txt, July 2005.

[BT02] BU, T. und D. TOWSLEY: *On Distinguishing between Internet Power Law Topology Generators*. In: *Proceedings of IEEE Infocomm*, June 2002.

[BZB+97] BRADEN, R., L. ZHANG et al.: *Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification*. RFC 2205, IETF, September 1997.

[CCG+02] CHEN, Q., H. CHANG et al.: *The Origin of Power Laws in Internet Topologies Revisited*. In: *Proceedings of IEEE Infocomm*, June 2002.

[CDK+05] CHANDRASHEKAR, J., Z. DUAN et al.: *Limiting Path Exploration in BGP*. In: *Proceedings of IEEE Infocom*, June 2005.

[CJ03] CRISTALLO, G. und C. JACQUENET: *Providing Quality of Service Indication by the BGP-4 Protocol: the QoS_NLRI attribute*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-jacquenet-qos-nlri-05.txt, Juni 2003.

[CR05] CAESAR, M. und J. REXFORD: *BGP Routing Policies in ISP Networks*. IEEE Network, 19(6):5–11, November 2005.

[CS02] CHANDRA, R. und J. SCUDDE: *Capabilities Advertisement with BGP-4*. RFC 3392, IETF, November 2002.

[CSG+06] CHARZINSKI, J., T. SCHWABE et al.: *Netzzugangssteuerung mit Toleranz für Inter-Domain Ausfälle*. Europäisches Patentamt, März 2006. Anmeldenummer: 10 2006 015 239.5.

[DCB+02] DAVIE, B., A. CHARNY et al.: *An Expedited Forwarding PHB (Per-Hob Behavior)*. RFC 3246, IETF, März 2002.

[DCK+04] DUAN, Z., J. CHANDRASHEKAR et al.: *Damping BGP Route Flaps*. In: *Proceedings of the 23rd IEEE International Performance Computing and Communications Conference (IPCCC)*, April 2004.

[DG05] DREXL, N. und C. GRUBER: *heuristics*. http://sourceforge.net/projects/heuristics/, 2005.

[DGG+99] DUFFIELD, N., P. GOYAL et al.: *A Flexible Model for Resource Management in Virtual Private Networks*. ACM SIGCOMM Computer Communication Review, 29(4):95–108, Oktober 1999.

[DGG+02] DUFFIELD, N., P. GOYAL et al.: *Resource Management With Hose: Point-to-Cloud Services for Virtual Private Networks*. IEEE/ACM Transaction on Networking, 10(5):679–692, Oktober 2002.

[Ebe06] EBERSPÄCHER, J.: *Entwicklungslinien und Trends des Internets*. Elektrotechnik & Informationstechnik, 6/2006:205–209, 2006.

[EGH+03] ENGEL, T., H. GRANZER et al.: *AQUILA: Adaptivr Resource Control for QoS Using an IP-based Layered Architecture*. IEEE Communications Magazine, 1, Januar 2003.

[Ent06] ENTHRONE PROJECT. http://www.enthrone.org, 2006.

[ESL+05] ENGEL, T., T. SCHWABE et al.: *Increasing End-to-End Availability over Multiple Autonomous Systems*. In: *Proceedings of the 2005 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA)*, June 2005.

[FFF99] FALOUTSOS, M., P. FALOUTSOS und C. FALOUTSOS: *On Power-Law Relationships of the Internet Topology*. In: *Proceedings of ACM Sigmetrics*, August 1999.

[FGL+01] FELDMANN, A., A. GREENBERG et al.: *Deriving Traffic Demands for Operational IP Networks: Methodology and Experience*. IEEE/ACM Transactions on Networking, 9:265–279, Juni 2001.

[FH99] FERGUSON, P. und G. HUSTON: *Quality of Service*. John Wiley and Sons, Inc., 1999.

[FKM+04] FELDMANN, A., H. KONG et al.: *Measuring BGP Pass-Through Times*. In: *Proceedings of the 5th Passive and Active Measurement Workshop (PAM)*, April 2004.

[FLY+93] FULLER, V., T. LI et al.: *Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy*. RFC 1519, IETF, September 1993.

[FSB+05] FU, X., H. SCHULZRINNE et al.: *NSIS: A New Extensible IP Signaling Protocol Suite*. IEEE Communications Magazine, Seiten 133–141, Oktober 2005.

[FT00] FORTZ, B. und M. THORUP: *Internet traffic engineering by optimizing OSPF weights*. In: *Proceddings of the IEEE INFOCOM*, March 2000.

[Gao01] GAO, L.: *On Inferring Autonomous System Relationships in the Internet*. IEEE/ACM Transactions on Networking, 9(6):733–745, December 2001.

[Gla03] GLASMANN, J.: *Ressourcenmanagement für Echtzeitverkehre in Intranets*. Dissertation, TU München, Institute of Communication Networks, 2003.

[GLH+00] GLEESON, B., A. LIN et al.: *A Framework for IP Based Virtual Private Networks*. RFC 2764, IETF, Februar 2000.

[Gol89] GOLDBERG, D.: *Genetic Algorithms in Search, Optimazation and Machine Learning*. Addison-wesley Publishing Company, Inc., Januar 1989.

[GP01] GRIFFIN, T. und B. PREMORE: *An Experimental Analysis of BGP Convergence Time*. In: *Proceedings of the 9th International Conference on Network Protocols (ICNP)*, November 2001.

[GR01] GAO, L. und J. REXFORD: *Stable Internet routing without global coordination*. IEEE/ACM Transactions on Networking, 9(6):681–692, Dezember 2001.

[Gri04] GRIFFIN, T.: *Bgp Wedgies*. In: *Proceedings of the Inter-domain Routing Workshop (IDRWS)*, May 2004.

[Gru06] GRUBER, C.: *GRAPH*. http://www.layonics.com/, 2006.

[GS05] GRIFFIN, T. und J. SOBRINHO: *Metarouting*. In: *Proceedings of the ACM SIGCOMM 2005*, August 2005.

[GSW02] GRIFFIN, T., F. SHEPHERD und G. WILFONG: *The Stable Paths Problem and Interdomain Routing*. IEEE/ACM Transactions on Networking, 10(2):232–243, April 2002.

[Hal97] HALABI, B.: *Internet Routing Architectures*. Cisco Press, 1997.

[HBW+99] HEINANEN, J., F. BAKER et al.: *Assured Forwarding PHB Group*. RFC 2597, IETF, Juni 1999.

[HCG06] HANSEN, A., T. CICIC und S. GJESSING: *Alternative Schemes for Proactive IP Recovery*. In: *Proceedings of the 2nd EuroNGI Conference on Next Generation Internet Design and Engineering*, April 2006.

[Hed88] HEDRICK, C.: *Routing Information Protocol*. RFC 1058, IETF, 1988.

[HFP+05] HOWARTH, M., P. FLEGKAS et al.: *Provisioning for Inter-domain quality of service: the MESCAL approach*. IEEE Communications Magazine, 43(6):129–137, Juni 2005.

[HHW+05] HO, K., M. HOWARTH et al.: *Two Approaches to Internet Traffic Engineering for End-to-end Quality of Service Provisioning*. In: *Proceedings of the 1st EuroNGI Conference on Next Generation Internet Networks*, April 2005.

[HKL+05] HANCOCK, R., G. KARAGIANNIS et al.: *Next Steps in Signaling (NSIS): Framework*. RFC 4080, IETF, Juni 2005.

[HPH+04] HO, K., G. PAVLOU et al.: *An incentive-based Quality of Service Aware Algorithm for Offline Inter-AS Traffic Engineering*. In: *Proceeding of the 2004 IEEE International Workshop on IP Operations and Management (IPOM)*, Oktober 2004.

[IET06] IETF WORKING GROUP: *NSIS, Next Steps in Signaling Working Group Charter*. http://www.ietf.org/html.charters/nsis-charter.html, 2006.

[IRY02] ITALIANO, G., R. RASTOGI und B. YENER: *Restoration Algorithm for Virtual Private Networks in the Hose Model*. In: *Proceedings of IEEE INFOCOM*, 2002.

[IT93] ITU-T: *Terms and Definitions Related Quality of Service and Network Performance Including Dependability*. Recommendation E.800, August 1993.

[IT04a] ITU-T: *General Overview of NGN*. Recommendation Y.2001, Dezember 2004.

[IT04b] ITU-T: *Gerneral principles and Gerneral Refernce Model for Next Generation Networks*. Recommendation Y.2011, Oktober 2004.

[JCJ00] JIN, C., Q. CHEN und S. JAMIN: *Inet: Internet Topology Generator*. Technischer Bericht CSE-TR-433-00, University of Michigan, 2000.

[JR05] JAGGARD, A. und V. RAMACHANDRAN: *Toward the Design of Robust Interdomain Routing Protocols*. IEEE Network, 19(6):35–41, November 2005.

[JRT04] JAISWAL, S., A. ROSENBERG und D. TOWSLEY: *Comparing the Structure of Power-Law Graphs and the Internet AS Graph*. In: *12th IEEE International Conference on Network Protocols (ICNP)*, 2004.

[KRS+01] KUMAR, A., R. RASTOGI et al.: *Algorithms for Provisioning Virtual Private Networks in the Hose Model*. In: *Proceedings of ACM SIGCOMM*, 2001.

[KW05] KATZ, D. und D. WARD: *Bidirectional Forwarding Detection*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-ietf-bfd-base-04.txt, Oktober 2005.

[LAB+00] LABOVITZ, C., A. AHUJA et al.: *Delayed Internet Routing Convergence*. In: *Proceedings of ACM Sigcomm*, August 2000.

[LAB+01] LABOVITZ, C., A. AHUJA et al.: *Delayed Internet routing convergence*. IEEE/ACM Transaction on Networking, 9(3):293–306, Juni 2001.

[LBM+04] LEVIS, P., M. BOUCADAIR et al.: *The Meta-QoS-Class concept: a step towards global QoS inter-domain services*. In: *Proceedings of the IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2004)*, Oktober 2004.

[LK05] LEE, C.-S. und D. KNIGHT: *Realization of the Next-Generation Network*. IEEE Communications Magazine, Seiten 34–41, Oktober 2005.

[Mae03] MAENNEL, O.: *Observed Properties of BGP Convergence*. Talk, Inter-Domain Routing Workshop (IDRWS), September 2003.

[MBA+04] MORAND, P., M. BOUCADAIR et al.: *Issues in MESCAL Inter-Domain QoS Delivery: Technologies, Bi-directionality, Inter-operability, and Financial Settlements*. Deliverable 1.4, MESCAL, http://www.mescal.org, Januar 2004.

[MBC+05] MORAND, P., M. BOUCADAIR et al.: *Final specification of protocols and algorithms for inter-domain SLS management and traffic engineering for QoS-based IP service delivery*. Deliverable 1.3, MESCAL, http://www.mescal.org, Juni 2005.

[MC05] MUSUNURI, R. und J. COBB: *An Overview of Solutions to Avoid Persistent BGP Divergence*. IEEE Network, 19(6):28–34, November 2005.

[Men04] MENTH, M.: *Efficient Admission Control and Routing for Resilient Communication Networks*. Dissertation, University of Würzburg, 2004.

[MGV+02] MAO, Z., R. GOVINDAN et al.: *Route Flap Damping Exacerbates Internet Routing Convergence*. In: *Proceedings of ACM Sigcomm*, August 2002.

[MKM06] MANNER, J., G. KARAGIANNIS und A. MCDONALD: *NSLP for Quality-of-Service Signaling*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-ietf-nsis-qos-nslp-10.txt, März 2006.

[MLM+01] MEDINA, A., A. LAKHINA et al.: *BRITE: An Approach to Universal Topology Generation*. In: *In Proceedings of the International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems- MASCOTS*, August 2001.

[MMB00] MEDINA, A., I. MATTA und J. BYERS: *On the origin of power laws in Internet topologies*. ACM Computer Communication Review, 30(2):18–28, April 2000.

[Moy98] MOY, J.: *OSPF Version 2*. RFC 2328, IETF, April 1998.

[MP02] MIHAIL, M. und C. PAPADIMITRIOU: *On the Eigenvalue Power Law*. In: *Proceedings of the 6th Intl. Workshop on Randomization and Approximation Techniques (RANDOM 2002)*, 2002.

[MVW+00] MUSUVATHI, M., S. VENKATACHARY et al.: *BGP-CT: A First Step Towards Fast Internet Fail-Over*. Technischer Bericht, Microsoft Research, 2000.

[MWA02] MAHAJAN, R., D. WETHERALL und T. ANDERSON: *Understanding BGP Misconfiguration*. In: *Proceedings of ACM Sigcomm 2002*, August 2002.

[Myk03] MYKONIATI, E.: *Admission Control for Providing QoS in DiffServ IP Networks. The TEQUILA Approach*. IEEE Communications Magazine, Seiten 38–44, Januar 2003.

[NBB+98] NICHOLS, K., S. BLAKE et al.: *Definition of the Differentiated Service Field (DS Field) in the IPv4 and IPv6 Headers*. RFC 2474, IETF, Dezember 1998.

[NC01] NICHOLS, K. und B. CARPENTER: *Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification*. RFC 3086, IETF, April 2001.

[NCM02] NYKVIST, J. und L. CARR-MOTYCKOVA: *Simulating Convergence Properties of BGP*. In: *Proceedings of the 11th International Conference on Computer Communications and Networks (IC3N)*, Seiten 124–129, Oktober 2002.

[Ora90] ORANROSEN, D.: *OSI IS-IS Intra-domain Routing Protocol*. RFC 1142, IETF, February 1990.

[PAM+98] PAXSON, V., G. ALMES et al.: *Framework for IP Performance Metrics*. RFC 2330, IETF, Mai 1998.

[PAM+05] PEI, D., M. AZUMA et al.: *BGP-RCN: Improving BGPConvergence through Root Cause Notification*. Computer Networks Journal, 48(2):175–194, Juni 2005.

[PHS00] PAN, P., E. HAHNE und H. SCHULZRINNE: *BGRP: Sink-Tree-Based Aggregation for Inter-Domain Reservations*. Journal of Communications and Networks, 2(2):157–167, Juni 2000.

[PMR+06] PSENAK, P., S. MIRTORABI et al.: *Multi-Topology (MT) Routing in OSPF*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-ietf-ospf-mt-06.txt, Februar 2006.

[Por06] PORETSKY, S.: *Considerations for Benchmarking IGP Data Plane Route Convergence*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-ietf-bmwg-igp-dataplane-conv-app-10.txt, März 2006.

[Pos81a] POSTEL, J.: *Internet Protocol*. RFC 791, STD 5, IETF, September 1981.

[Pos81b] POSTEL, J.: *Transmission Control Protocol*. RFC 793, IETF, September 1981.

[Pre03] PREMORRE, B.: *An Analysis of Convergence Properties of the Border Gateway Protocol using Discrete Event Simulation*. PhD thesis, Dartmouth College Hanover, New Hampshire, 2003.

[PSS05] PRZYGIENDA, T., N. SHEN und N. SHETH: *M-ISIS: Multi Topology (MT) Routing in IS-IS*. Internet Draft, http://www.ietf.org/internet-drafts/draft-ietf-isis-wg-multi-topology-11.txt, Oktober 2005.

[RBF03] RIEDL, A., T. BAUSCHERT und J. FRINGS: *On the Dimensioning of Voice over IP Networks for various Call Admission Control Schemes*. In: *Proceedings of the 18th International Teletraffic Congress (ITC 18)*, August 2003.

[rib06] *The RIBE Network Coordination Centre*. http://www.ripe.net, 2006.

[Rie02] RIEDL, A.: *A Hybrid Genetic Algorithm for Routing Optimization in IP Networks Utilizing Bandwidth and Delay Metrics*. In: *Proceedings of the IEEE Workshop on IP Operations and Management (IPOM)*, October 2002.

[Rie04] RIEDL, A.: *Routing Optimization and Capacity Assignment in Multi-Service IP Networks*. Dissertation, TU München, Institute of Communication Networks, 2004.

[RLH06] REKHTER, Y., T. LI und S. HARES: *A Border Gateway Protocol 4 (BGP-4)*. RFC 4271, IETF, Januar 2006.

[RM04] REICHERT, C. und T. MAGEDANZ: *A Fast Heuristic for Genetic Algorithms in Link Weight Optimization*. In: *Proceedings of QofIS*, September 2004.

[RMD05] RAI, S., B. MUKHERJEE und O. DESHPANDE: *IP Resilience within an Autonomous system: Current Approaches, Challenges, and Future Directions*. IEEE Communications Magazine, Seiten 142–149, Oktober 2005.

[rou05] *Oregon Route Views Project*. http://www.antc.uoregon.edu/route-views, 2005.

[RRT06] RATHGEB, E., S. RIEBACH und B. TÖDTMANN: *Neue Herausforderungen für den Schutz kritischer Infrastrukturkomponenten in zukünftigen IP-Netzen*. Elektrotechnik & Informationstechnik, 6/2006:215–223, 2006.

[RVC01] ROSEN, E., A. VISWANATHAN und R. CALLON: *Multiprotocol Label Switching Architecture*. RFC 3031, IETF, Januar 2001.

[Sch06a] SCHWABE, T.: *Achieving the Minimal BGP Convergence Time*. In: *Proceedings of the 2nd EuroNGI Conference on Next Generation Internet Design and Engineering*. EuroNGI, April 2006.

[Sch06b] SCHWABE, T.: *Considering BGP Routing for the Traffic Engineering of uncertain Traffic Demands*. In: *Proccedings of the Networks 2006*. VDE, IEEE, November 2006.

[SE04] SCHWABE, T. und T. ENGEL: *Independence of Inter-domain QoS Signaling and Routing*. In: *Proceedings of the Inter-domain Routing Workshop (IDRWS)*, May 2004.

[SG05] SCHWABE, T. und C. GRUBER: *Traffic Variations Caused by Inter-domain Rerouting*. In: *Proceedings of the 5th International Workshop on Design of Reliable Communication Networks (DRCN)*, Oktober 2005.

[SGV03] SOFIA, R., R. GUERIN und P. VEIGA: *SICAP, A Shared-segment Inter-domain Control Aggregation Protocol*. In: *Proceedings of the Workshop on High Performance Switching and Routing 2003 (HPSR)*, June 2003.

[SH06] SCHULZRINNE, H. und R. HANCOCK: *GIST: General Internet Signaling Transport*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-ietf-nsis-ntlp-10.txt, Juli 2006.

[Sha05a] SHAND, M.: *IP Fast Reroute Framework*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-ietf-rtgwg-ipfrr-framework-04.txt, 2005.

[Sha05b] SHARAFEDDINE, S.: *IP Network Planning for Realtime Services with Statistical QoS Guarantees*. Dissertation, TU München, Institute of Communication Networks, 2005.

[Sof03] SOFIA, R.: *Shared-segment Inter-domain Control Aggregation Protocol*. Doktorarbeit, University of Lisbon, Oktober 2003.

[SPG97] SHENKER, S., C. PARTRIDGE und R. GUERIN: *Specification of Guaranteed Quality of Service*. RFC 2212, IETF, September 1997.

[ssf06] *Scalable Simulation Framework Network Models (SSFNet)*. http://www.ssfnet.org/homePage.html, 2006.

[SSL+05] SHEN, C., H. SCHULZRINNE et al.: *Routing Dynamics Measurement and Detection for Next Step Internet Signaling Protocol*. In: *Proceedings of the IEEE/IFIP Workshop on End-to-End Monitoring Techniques and Services (e2emon)*, Mai 2005.

[Tel05] TELSTRA: *AS1221 BGP Table Data*. http://bgp.potaroo.net/as1221/bgp-active.html, 2005.

[Tro05] TROMPARENT, M.: *Resource Management for Real-time Traffic - Fault-tolerance, Scalability and Mobility Support*. Dissertation, TU München, Institute of Communication Networks, 2005.

[TSG+04] TEIXEIRA, R., A. SHAIKH et al.: *Dynamics of Hot-Potato Routing in IP Networks*. In: *Proceedings of ACM Sigmetrics 2004*, June 2004.

[UBQ03] UHLIG, S., O. BONAVENTURE und B. QUOITIN: *Interdomain Traffic Engineering with minimal BGP Configurations*. In: *Proceedings of the 18th International Teletraffic Congress (ITC)*, September 2003.

[Uhl04] UHLIG, S.: *Implications of the traffic characteristics on interdomain traffic engineering*. PhD thesis, Université catholique de Louvain, 2004.

[VCG98] VILLAMIZAR, R., R. CHANDRA und R. GOVINDAN: *BGP Route Flap Damping*. RFC 2439, IETF, November 1998.

[VCP+03] VERBRUGGE, S., D. COLLE et al.: *Common planning practices for network dimensioning under traffic uncertainty*. In: *Proceedings of the 4th International Workshop on the Design of Reliable Communication Networks (DRCN)*, Oktober 2003.

[WMR+05] WU, J., Z. MAO et al.: *Finding a needle in a haystack: Pinpointing significant BGP routing changes in an IP network,*. In: *Proceedings of 2nd symposium on Networked Systems Design and Implementation (NSDI)*, May 2005.

[WRC06] WALTON, D., A. RETANA und E. CHEN: *Advertisement of Multiple Paths in BGP*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-walton-bgp-add-paths-05.txt, Februar 2006.

[Wro97] WROCLAWSKI, J.: *Specification of the Conrolled-Load Network Element Service*. RFC 2211, IETF, September 1997.

[WS98] WATTS, D. und S. STROGATZ: *Collective dynamics of 'small-world' networks*. Nature, 393:440–442, 1998.

[YFMB+04] YANNUZZI, M., A. FONTE et al.: *A proposal for inter-domain QoS routing based on distributed overlay entities and QBGP*. In: *Proceedings of the First International Workshop on QoS Routing (WQOSR)*, October 2004.

[YFMB+05] YANNUZZI, M., A. FONTE et al.: *A Self-Adaptive QoS Routing Framework for Multi-homed Stub Autonomous Systems*. In: *Proceedings of the E-NEXT Doctoral Summer School (EUNICE)*, July 2005.

[YMBB05] YANNUZZI, M., X. MASIP-BRUIN und O. BONAVENTURE: *Open Issues in Interdomain Routing: A Survey*. IEEE Network, 19(6):49–56, November 2005.

[ZAL03] ZHANG, H., A. ARORA und Z. LIU: *G-BGP: Stable and Fast Convergence of the Border Gateway Protocol*. Technischer Bericht OSU-CISRC-6/03-TR36, Ohio State University, Juni 2003.

[ZDL05] ZHAI, S., X. DUAN und L. LI: *BFD initializtion with BGP and static routes*. Internet Draft, IETF, http://www.ietf.org/internet-drafts/draft-suping-bfd-bgp-static-app-02.txt, Juni 2005.

[Zin01] ZININ, A.: *Cisco IP Routing*. Addison-Wesley, 2001.

[ZSN05] ZHAO, M., S. SMITH und D. NICOL: *The Performance Impact of BGP Security*. IEEE Network, 19(6):42–48, November 2005.

[Zus06] ZUSE INSTITUTE BERLIN: *SNDlib*. http://sndlib.zib.de/home.action, 2006.