

# Technik INFO



## Bandbreitenmanagement in Kommunikationsnetzen

Ein Arbeitspapier des VAF Bundesverband Telekommunikation e.V.



## Impressum und Nutzungshinweise

### Technik INFO

#### Bandbreitenmanagement in Kommunikationsnetzen

Ein Arbeitspapier des VAF Bundesverband Telekommunikation e.V.

#### Stand

Mai 2010

#### Autoren

Mathias Hein und Prof. Dr. Bernhard Stütz

#### Fachliche Mitwirkung

In den vorliegenden Text sind zahlreiche fachliche Hinweise, inhaltliche Anregungen und Kommentare eingeflossen. Stellvertretend seien in alphabetischer Namensreihenfolge genannt:

Frank Bauer (Aastra Deutschland GmbH)

Jürgen Brieskorn (Siemens Enterprise Communications GmbH & Co.KG)

Dr. Jörg Fischer (Alcatel-Lucent Deutschland AG)

Olaf Hagemann (Extreme Networks GmbH)

Holger Häntzschel (Computacenter AG & Co.oHG)

Michael Herwig (Telefonbau Schneider GmbH & Co.KG)

Benjamin Kolbe (Nextragen GmbH)

Rainer Wiener (Avaya Deutschland GmbH)

Ludwig Wurmstatter (Microsoft Deutschland GmbH)

Das Arbeitspapier basiert auf einer Textvorlage der Autoren aus dem Frühjahr 2009 für den Fachkreis Netzwerktechnik im VAF. Dank gilt allen Kommentatoren, die zur Entwicklung des Textes beigetragen haben. Die inhaltliche Verantwortung verbleibt bei den Autoren, durch welche die zum aktuellen Stand finale Textfassung erstellt wurde.

#### Herausgeber

VAF Bundesverband Telekommunikation e.V.

Otto-Hahn-Str. 16, 40721 Hilden

[www.vaf-ev.de](http://www.vaf-ev.de)

#### Layout / Gestaltung

Uwe Klenner, Layout und Gestaltung,

Rittsteiger Str. 104, 94036 Passau

[www.layout-und-gestaltung.de](http://www.layout-und-gestaltung.de)

#### Copyright: VAF 2010

Alle Rechte, auch das der auszugsweisen Vervielfältigung, liegen beim VAF.

## Inhaltsverzeichnis

<b>Impressum und Nutzungshinweise</b> .....	<b>S. 2</b>	<b>4. Quality-of-Service-Techniken</b> .....	<b>S. 29</b>
<b>Vorbemerkung</b> .....	<b>S. 4</b>	4.1 Virtuelle LANs .....	S. 30
<b>Bearbeitungshinweise</b> .....	<b>S. 6</b>	4.2 IEEE 802.1p/Q .....	S. 31
<b>1. Sprachqualität in Netzwerken</b> .....	<b>S. 7</b>	4.3 Integrated-Services .....	S. 33
1.1 Subjektive und objektive Bewertung		4.4 Differentiated-Services .....	S. 37
der Sprachqualität .....	S. 9	4.4.1 Exkurs zu DiffServ:	
1.2 Verzögerung .....	S. 14	Verwurfregeln bei Überlast .....	S. 42
1.3 Paketverlustrate .....	S. 17	4.5 Traffic-Shaping .....	S. 46
1.4 Jitter .....	S. 18	4.6 Queueing .....	S. 47
1.5 Datendurchsatz .....	S. 20	4.7 Sicherung der Dienstgüte im WAN .....	S. 51
<b>2. Sprachqualität in IP-Netzen</b> .....	<b>S. 21</b>	4.8 Stolpersteine im Bandbreitenmanagement .....	S. 52
2.1 Network-Service-Quality .....	S. 23	4.9 Zur Diskussion: DiffServ oder IEEE 802.1p/Q? .....	S. 54
2.2 Effektive Bandbreite .....	S. 23	<b>5. Quality-of-Experience</b> .....	<b>S. 55</b>
2.3 Bandbreitenmanagement .....	S. 24	5.1 Breitbandige Sprache zur Verbesserung	
<b>3. Überdimensionierung und Rightsizing</b> .....	<b>S. 26</b>	der Telefonlösungen .....	S. 57
3.1 Überdimensionierung im WAN .....	S. 27	5.2 Intelligente, adaptive Endpunkte .....	S. 57
3.2 Rightsizing .....	S. 27	5.3 Messungen und Monitoring .....	S. 58
		5.4 Mediastack .....	S. 58
		<b>6. Fazit</b> .....	<b>S. 62</b>
		<b>Schlussbemerkung</b> .....	<b>S. 64</b>
		<b>Literatur und Weblinks</b> .....	<b>S. 65</b>

## Vorbemerkung

### Warum dieses Arbeitspapier?

Verfügbarkeit, Sprachqualität und Zuverlässigkeit des Dienstes Telefonie sind für den geschäftlichen Nutzer nicht die einzigen, aber die elementarsten Anforderungen im täglichen Betrieb. Durch den Einsatz der VoIP-Technologie in Unternehmen verlagern sich wesentliche Mechanismen zur Steuerung der Dienstqualität von einer separierten TK-Anlage mit exklusivem Netz in die LAN/WAN-Infrastruktur des Unternehmens. Die Telefonie als Echtzeit-Dienst teilt sich dabei die effektive Bandbreite mit sämtlichen Diensten des Datennetzes. Der Einsatz von VoIP kann eine Reihe von Vorteilen mit sich bringen. Hierzu gehört insbesondere die Effizienzsteigerung der Kommunikationsprozesse durch VoIP als Teil von umfassenden Lösungen für »Unified Communications & Collaboration«.

Zugleich verwandelt VoIP das in Bezug auf Übertragungszeit tolerante Datennetz in ein Kommunikationsnetz, welches zusätzlich und durchgängig die spezifischen Anforderungen der zeitkritischen Anwendung Sprache mit bedienen muss. Diese Einsatzerweiterung gelingt nur, wenn das Netz für die erweiterten Aufgabenstellungen angepasst wird. Empfängt der Anwender eine E-Mail eine Sekunde verzögert, merkt er dies nicht.

Ein verspäteter, um Sekundenbruchteile verzögerter Empfang von Sprachdaten sabotiert dagegen den Dienst »Voice«.<sup>1</sup>

Insellösungen sind im laufenden Betrieb regelmäßig zum Scheitern verurteilt. Gefordert ist eine ganzheitliche Betrachtungsweise. Im Zentrum steht die technisch und wirtschaftlich optimale Bereitstellung der für den Dienst Sprache benötigten Bandbreite. Bandbreite alleine ist zu jedoch wenig. Es geht um die komplette Realisierung von Quality-of-Service (QoS), d.h. um eine entsprechend bevorzugte Behandlung der multimedialen Datenströme. Echtzeit bedeutet auch und vor allem die Signalisierung in Echtzeit.

Für die Aufgabe des Bandbreitenmanagements in Kommunikationsnetzen existieren unterschiedliche, zum Teil konkurrierende, zum Teil redundante Methoden.

### Was liefert das Arbeitspapier?

Dieses Arbeitspapier bietet dem Leser einen Überblick über die wesentlichen und in aktuellen Diskussionen vertretenen Verfahren, legt die Beziehung der netzwerkorientierten QoS-Techniken sowie der Ansätze zu deren Erweiterung in den End-

1) Die in diesem Arbeitspapier aufgeführten Qualitätsanforderungen finden grundsätzlich auch für die Übermittlung von Video über IP-Netzwerke Anwendung.



gerätebereich dar und ergänzt die Darstellung um praktische Hinweise. Behandelt werden:

- ▶ Netzdimensionierung (Overprovisioning und Rightsizing),
- ▶ Verkehrsklassen-Klassifizierung (Priorisierung),
- ▶ Signalisierungs-Mechanismen und -Verfahren,
- ▶ intelligente Warteschlangenverwaltung (Queuing und Scheduling),
- ▶ intelligente Bandbreitenverwaltung (Traffic-Policing),
- ▶ Bandbreitenzuweisung (Admission-Control) sowie
- ▶ Quality-of-Experience.

Ausgangspunkt für jede Überlegung zum Bandbreitenmanagement in Kommunikationsnetzen ist die Sprachqualität. Aus diesem Grund beginnt dieses Arbeitspapier mit einer Darstellung der Grundbegriffe, technischen Eigenschaften und Bewer-

tungsmethoden der Sprachqualität. Auch werden dort die wesentlichen Unterschiede zwischen klassischem Telefon- und Datennetz nochmals in kompakter Form dargelegt. Im Fazit stehen die Zusammenschau sowie praktische Hinweise für das technische Management im Vordergrund.

Naturgemäß kann ein solches Arbeitspapier zu aktuellen Standards, anerkannten Regeln der Technik, praktischen Erfahrungen und Innovationen keine Abhandlung sein, die gleichsam in Stein gemeißelt ist. Der interessierte Leser ist darum eingeladen, durch Anregungen und Hinweise, aber auch durch Fragen zur Fachdiskussion beizutragen.

**Martin Bürstenbinder**, Geschäftsführer,  
VAF Bundesverband Telekommunikation e.V.

## Bearbeitungshinweise zum Arbeitspapier

- ▶ Das Arbeitspapier basiert auf einer Textvorlage der Autoren aus dem Frühjahr 2009 für den Fachkreis Netzwerktechnik im VAF. Dank gilt allen Kommentatoren, die zur Entwicklung des Textes beigetragen haben. Die Auswertung und Erstellung der abschließenden Textfassung und damit auch die inhaltliche Verantwortung liegt bei den Autoren.
- ▶ Nicht aufgenommen wurden die sachlich berechtigten Hinweise auf die Bedeutung der akustischen und damit auch baulichen Eigenschaften von Endgeräten für die resultierende Sprachqualität in einer Ende-zu-Ende-Betrachtung. Die angemessene Berücksichtigung der damit verbundenen Aspekte würde den Rahmen dieser Abhandlung sprengen und bedarf einer gesonderten Befassung.
- ▶ Das ursprüngliche Kapitel zu Integrated-Services wurde zu einem Unterkapitel zusammenge Kürzt. Trotz mancher Empfehlung, dieses Thema – mangels praktischer Relevanz insbesondere im Bereich Privater Netze – völlig auszusparen, wurde die verkürzte Fassung beibehalten. Das Arbeitspapier will eine kompakte Basis für Fachdiskussionen liefern, in diesen tritt Integrated-Services unverändert in Erscheinung und wird gerne als die »eigentlich bessere Lösung« ins Feld geführt. Das kann und soll nicht ignoriert werden.
- ▶ Alle Ausführungen in den Kapiteln 1 bis 4 sind streng der Produktneutralität verpflichtet, auf internationale Standards hin orientiert und schöpfen aus einem umfangreichen Pool an Erfahrungen und Expertenwissen. Hinweise auf Hersteller im Literaturverzeichnis sind nicht als Einschränkung zu verstehen, sondern dienen dem Quellenhinweis für die Überlassung

von technischen Angaben, deren produktneutrale Geltung von den Autoren bestätigt wird.

- ▶ Zusätzlich soll hier darauf hingewiesen werden, dass unterschiedlichste Hersteller teilweise über die etablierten Standards hinaus, teilweise parallel dazu und manchmal auch aus Ermangelung an etablierten Standards Mechanismen und Implementierungen bieten, die in homogenen oder im jeweiligen Einzelfall dafür zertifizierten Umgebungen des Herstellers der Leistungsverbesserung dienen sollen (sogenannte proprietäre Lösungen). Da diese jedoch nicht in jedem Fall oder nicht ohne Einschränkungen oder Einzelfallprüfung allgemein übertragbar sind, beziehungsweise in heterogenen Systemumgebungen nicht unbedingt vollständig interoperabel oder bei Erweiterungen mit standardkonformen Produkten von Drittherstellern fallweise überhaupt nicht anwendbar sein können, werden sie hier nicht erwähnt.

- ▶ Ausnahme zu dem Neutralitätsgrundsatz bildet das Kapitel 5 zu Quality-of-Experience. Hier wird unter wesentlicher Bezugnahme auf Herstellerangaben ein Ansatz vorgestellt, der – nicht in jedem einzelnen Element, jedoch in einigen wesentlichen und schließlich als Ganzes gesehen – spezifisch von dem Hersteller Microsoft stammt und dementsprechend proprietär funktioniert. Gemäß der schon erwähnten Aufgabenstellung dieses Arbeitspapiers, einen Beitrag für weitere Fachdiskussionen zu aktuell gestellten Fragen an die Hand zu geben, sind die Autoren und der Herausgeber der Auffassung, dass dieses Vorgehen zweckmäßig ist. Technische Wertungen sind damit nicht verbunden.

## 1. Sprachqualität in Netzwerken

Die Übertragung von Telefongesprächen über Ethernet- und IP-basierte Datennetze stellt an diese neue Anforderungen. Wie gut die Sprachqualität in Netzwerken ist, ermitteln verschiedene Verfahren.

Das klassische Telefonnetz baut auf der Technik des Time-Division-Multiplexing (TDM) zur Übermittlung der digitalisierten Sprachinformationen auf. Dieser Mechanismus sorgt dafür, dass in einem Übertragungskanal beziehungsweise über eine Übertragungsstrecke zwei oder mehrere Signale beziehungsweise Signalströme zeitlich verschachtelt übertragen werden können. Hierzu erhält jeder Unterkanal einen definierten Zeitschlitz fester Länge. Da die Übertragungsressourcen zeitlich begrenzt sind, legt die TDM-Technik immer eine feste Anzahl von Sub- beziehungsweise Unterkanälen fest.

Die Subkanäle erhalten jeder eine konstante Bandbreite pro Zeiteinheit. Dadurch ist die Verzögerung zwischen den Kommunikationspartnern innerhalb eines Übertragungskanals konstant. Der gesamte Übertragungskanal zwischen Sender und Empfänger bleibt so lange bestehen, bis die Teilnehmer die Ressourcen wieder freigeben. Werden über einen bestehenden Kanal zwischen zwei Teilnehmern keine Informationen übermittelt, bleibt der im Leerlauf befindliche Übertragungskanal erhalten und die Ressourcen können nicht von anderen Teilnehmern genutzt werden. Der Vorteil des TDM-Verfahrens liegt in dem geringen Protokoll-Overhead und der optimalen Nutzung der verfügbaren Bandbreite.

### Datennetze sind für die klassische Datenübertragung ausgelegt

Bei paketvermittelten Netzen nutzen alle aktiven Teilnehmer die verfügbaren Übertragungsressourcen. Die Datenpakete erhalten spezifische Adressinformationen und können dadurch von den Knotenpunkten individuell vermittelt werden. Aus diesem Grund können Pakete zwischen einem Absender und einem Empfänger über unterschiedliche Netzwege übertragen werden. Früher wurden deshalb über Datennetze nur solche Daten übertragen, die keine Anforderungen an die Isochronität<sup>2</sup> stellten. Entscheidend war es, dass die Daten in einem konstanten Zeitraster verlustfrei, verzögerungsarm und so sicher wie möglich zwischen den beiden Endpunkten übertragen wurden.

Ein weiteres Merkmal der Datentransfers in Datennetzen ist die große Varianz in Bezug auf ihr Datenaufkommen. Wenn Lastspitzen entstehen, kommt es zu sogenannten Datenbursts. Rechner kommunizieren in kurzen Paketbursts, die so

2) Isochronität: Beschreibt die Eigenschaft eines technischen Systems, insbesondere eines Netzwerks oder anderen Übertragungssystems, definierte Aufgaben innerhalb einer genau definierten Zeitspanne durchzuführen. Isochrone Telekommunikationsnetzwerke können fehlerfrei mit konstanter Datenrate übertragen.



schnell wie möglich übertragen werden müssen, um dann wieder für mehrere Minuten die Kommunikation völlig einzustellen. Auch können im Gegensatz zu TDM bei IP die Pakete unterschiedlich groß sein. Die Paketlänge hat somit einen direkten Einfluss auf die Realisierung von QoS. Klassische Rechnernetze sind auf dieses Verhalten entsprechend angepasst. Dies gilt sowohl für lokale Netzwerke als auch für typische Weitverkehrsnetze.

### Neue Anforderungen entstehen durch Sprachübertragung

Für Multimedia-Anwendungen, beispielsweise Voice-over-IP oder Videokonferenzen, die unterbrechungsfreie, verlustarme

und verzögerungsarme Datenströme benötigen, ist diese Funktionsweise völlig unbrauchbar. Solche Dienste stellen eine Reihe von Anforderungen wie geringe Verzögerung oder garantierte konstante Übertragungsraten, die bisher in Netzen nicht bestanden. Die Übertragung von Informationen mit einer definierten Dienstgüte ist das zentrale Merkmal der Entwicklung besserer Datennetze. Jeder Mechanismus zur Sicherung der Dienstgüte muss sich an der Übertragungsqualität der Sprachnetze messen lassen.

Jede Anwendung erwartet von einem Netzwerk bestimmte Eigenschaften. Die Übertragung von Sprache in IP-Netzen erfordert eine garantierte Qualität, die mit der aus den Telefonnetzen vergleichbar sein muss. Erforderlich sind Kenntnisse über diejenigen Parameter, die die Qualität einer Übertragung bestimmen. Die Verarbeitungsprozesse in aktiven Komponenten wie Gateways, Switches oder Router sind die Quelle der meisten Beeinflussungen. Daher müssen neue Mechanismen diese Prozesse verbessern, um die Qualität einer Übertragung zu garantieren.

Für eine bidirektionale Kommunikation müssen beide Teilnehmer in der Lage sein, sowohl Informationen zu senden, als auch zu empfangen. Der Übertragungsweg kann näher spezifiziert werden, wenn das verwendete Trägermedium und die technischen Einrichtungen bekannt sind, also die Kommunikation über große Entfernungen möglich sein soll. Die Zielstellung, eine definierte Dienstgüte in Datennetzen einzuführen, verlangt Grundkenntnisse über Anforderungen an den Informationstransport und damit über die Art der übertragenen Information. Voice-over-IP, Videokonferenzen, Multimedia-over-IP, Multiservice-Networking, Real-Time-Networking,

#### Klassische Telefonnetze

sprachorientiert

---

garantiertes QoS

---

hohe Konzentration in den Vermittlungsstellen, die für eine mittlere Belegungsdauer von 3 Minuten konzipiert sind

---

wichtigster Vertreter ist das Telefonnetz als größtes Netz der Welt mit über 1,5 Mrd. Anwendern

---

geringes Wachstum

---

zentrale Intelligenz innerhalb des Netzes

---

strenge Standards, aber keine offenen Systeme, hohe Komplexität

---

hochspezialisierte Vermittlungs- und Protokollmaschinen

---

höchste Verfügbarkeit, Zuverlässigkeit und Reaktionszeit durch proprietäre Parallelrechner

#### Klassische Datennetze

datenorientiert

---

QoS abhängig von der Technologie und der zur Verfügung stehenden Bandbreite

---

nur von der Bandbreite im Übertragungsnetz abhängig

---

wichtigster Vertreter ist das Internet mit über 1 Mrd. Anwendern

---

extrem hohes Wachstum

---

verteilte Intelligenz

---

offene Systeme und Standards mit vielen Interpretationsmöglichkeiten

---

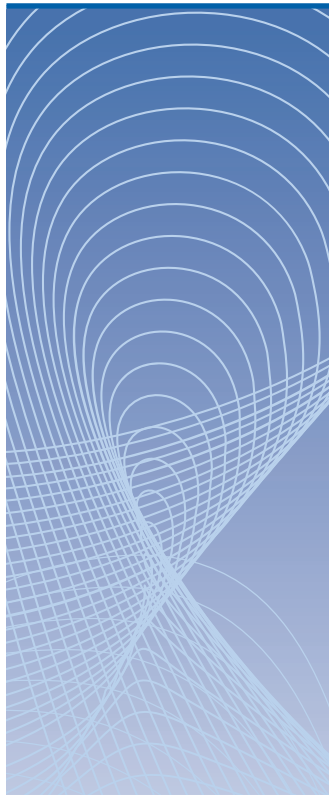
offene Systeme, Protokolle sind weltweit als Public Domain verfügbar

---

hohe Verfügbarkeit durch verteilte Netzintelligenz, hohe Flexibilität und Anpassungsfähigkeit

Vergleich klassischer Telefonnetze vs. klassischer Datennetze





MOS	Sprachqualität
5	ausgezeichnet
4	gut
3	ordentlich
2	mäßig
1	schlecht

MOS-Skala zur Bewertung der Sprachqualität

konvergente Netze – die Anwendungen, die auf einem Daten-netz laufen sollen, verändern sich rasant und stellen erhöhte Anforderungen an die Netzwerktechnologie, die alle Informa-tions- und Kommunikationsarten in einem IP-Netz vereint.

### Entscheidend ist die wahrgenommene Qualität

Die ultimative Messgröße zur Beurteilung der Performance ist der Anwender. Bei der Sprachübermittlung hängt die Beurteilung der Güte von Sprachübermittlungen von subjektiven Kri-terien ab. Es gibt viele Faktoren, die Einfluss darauf haben. Der offensichtlichste davon ist die Qualität der Mikrophone und Lautsprecher auf beiden Seiten. Die Digitalisierung und Kodierung des Signals ist ein weiterer wichtiger Faktor. Vor allem die Wahl des Kodierungsschemas ist entscheidend für den Rauschabstand und die Verzerrung des Signals. Wenn das Ko-dierungsschema auf der Strecke zwischen Sender und Emp-fänger zusätzlich verändert wird, leidet darunter die Qualität. Wird Silence-Suppression eingesetzt, haben auch die Güte des Comfort-Noise und eventuelles Clipping Einfluss. Auch ist der Einfluss des Telefonendgerätes (TEG) zu berücksichtigen. Hier sei lediglich darauf hingewiesen, dass die Gesamtausstattung des TEG' s zu berücksichtigen ist und eine der wesentlichen Funktionen moderner TEG' s die so genannte Akustik-Waage darstellt. Es kann nur das aus einem TEG herauskommen, was in das Gerät herein geht.

Die von den Benutzern erwartete Sprachqualität variiert stark und hängt von vielen Faktoren ab. Hierzu gehören beispiels-weise kulturelle Unterschiede, Anforderungen aus dem Ge-schäftsumfeld, Umgebungsbedingungen oder verfügbare Hard- und Software. Auch hängt die Toleranz gegenüber Stö-

rungen stark vom jeweiligen Kommunikationsgerät ab. Die To-leranzschwellen von Handy-Nutzern sind wesentlich höher als von Nutzern klassischer Telefondienste.

Die höchsten Anforderungen an die Sprachkommunikation stellt die Geschäftskommunikation. Aus diesem Grund werden die Voice-over-IP-Lösungen (VoIP) immer an den Merkmalen der klassischen ISDN-Telefonanlagen gemessen. Somit gehört die Sprachqualität zu den fundamentalen Kriterien jeder Sprachanwendung. Diese kann in Netzwerken – sowohl im LAN als auch im WAN – nicht problemlos garantiert werden und erfordert das reibungslose Zusammenspiel komplexer Tech-nologien.

## 1.1 Subjektive und objektive Bewertung der Sprachqualität

Die Sprachqualität beschreibt, wie gut die Verständlichkeit ei-ner menschlichen Stimme bei Aufzeichnung und Wiedergabe durch eine technische Einrichtung ist. Eine Beurteilung der Sprachqualität ist dabei subjektiv und hängt sowohl von den gegebenen technischen Mitteln, dem Umfeld der Aufnahme, dem Übertragungsweg und dem Umfeld der Wiedergabe ab. Die Bewertung dieser Sprachqualität ist durch Bewertungsme-thoden der ITU mit dem Standard P.800 spezifiziert.

### Subjektive Bewertung der Sprachqualität

Das bekannteste Verfahren zur Bewertung der Sprachqualität ist der »Mean Opinion Score«, kurz MOS. Die ITU hat die Tech-nik Mean Opinion Score im Jahre 1998 als Empfehlung P.800

veröffentlicht. Diese Methode beschreibt die subjektive Wahrnehmung der verschiedenen »Prüflinge«, z.B. Codecs mit Hilfe einer festgelegten Skala zur Beurteilung der QoS-Empfindungswerte. Der MOS-Wert ist ein dimensionsloser Wert ähnlich den Schulnoten zwischen eins und fünf. Dabei steht der Wert »1« für eine mangelhafte Sprachqualität, bei der keine Verständigung möglich ist, der Wert »5« hingegen signalisiert eine exzellente Übertragungsqualität, die nicht von dem Original zu unterscheiden ist.

Der MOS-Wert wird subjektiv ermittelt, indem Sprechproben Probanden vorgespielt, die einzelnen Bewertungen gewichtet und daraus die statistischen Ergebnisse ermittelt werden. Die einzelnen Schritte bei der MOS- Technik ergeben sich wie folgt:

1. Es wurde ein Publikum von mindestens 100 Menschen befragt.
2. Danach wurde eine Sprachprobe mit dem entsprechenden Sprachcodec kodiert und diese Sprachsequenz dem Publikum vorgespielt.
3. Jeder Einzelne vergab danach für die verminderte Qualität der Sprachprobe eine Note. Diese musste zwischen 1 (schlecht) und 5 (ausgezeichnet) liegen.
4. Aus den 100 Noten wurde der Durchschnitt berechnet und als MOS- Wert für den Sprachcodec festgelegt.

Die nachfolgende Tabelle zeigt die gängigsten Codecs und den ermittelten MOS- Wert. Die abgebildeten MOS- Werte entsprechen der besten Qualität, die ein Sprachcodec erhalten kann.

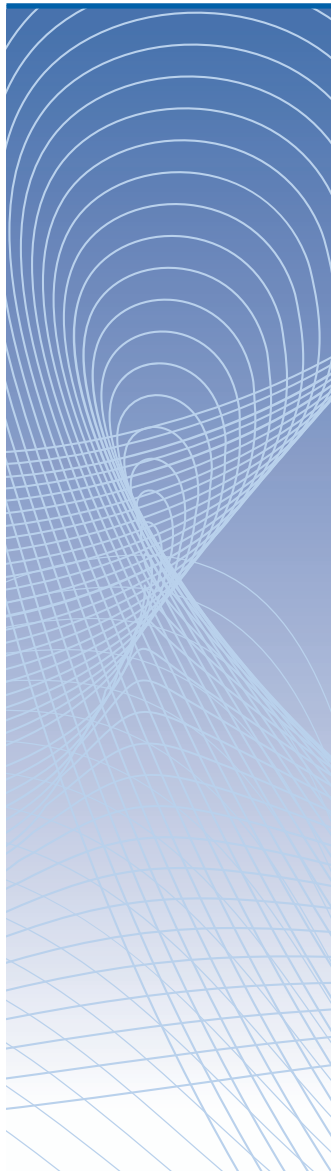
Codec	MOS
G.711	4,1
G.722.2	4,1
G.729	3,92
G.726	3,85
iLBC	3,8
G.729a	3,7
G.723.1	3,65
G.728	3,61

Tabelle: Codecs und deren dazugehörige MOS- Werte

MOS bedeutet in seiner Ursprungsform Mean Opinion Score, also durchschnittlicher Meinungswert. Er wird von vielen Verfahren und Algorithmen als Ausgabewert benutzt. Diese Werte lassen sich nicht immer genau vergleichen, da die unterschiedlichen Verfahren divergierende Ansätze haben. Deshalb hat die ITU in der Veröffentlichung Rec. P.800.1 beschrieben, dass dem Wort »MOS« noch Zeichen angehängt werden müssen, um eindeutig zu zeigen, welche Art von Verfahren dem berechneten MOS- Wert zu Grunde liegt.

In dieser Empfehlung werden folgende Zeichen erklärt, die an das Wort MOS anzuhängen sind:

- LQ** – Listen Quality (Gehörte Qualität)
- CQ** – Conversational Quality (Dialog- Qualität)
- S** – Subjective
- O** – Objective
- E** – Estimated (Geschätzt)



Die nachfolgende Tabelle zeigt die daraus resultierenden Kombinationen:

	Listening-only	Conversational
Subjektive	MOS- LQS	MOS- CQS
Objektive	MOS- LQO	MOS- CQO
Geschätzte	MOS- LQE	MOS- CQE

Tabelle: Verfügbare MOS- Terminologien

**MOS- LQS:** Dieser Wert wurde in einem Labor erhoben. Das arithmetische Mittel wird aus gesammelten subjektiven Beurteilungen berechnet, welche zwischen 1 und 5 liegen, wie in der Empfehlung Rec. P.800 beschrieben. Ergebnisse basierend auf der Empfehlung Rec. P.830 liefern MOS-LQS.

**MOS- LQO:** Dieser Wert soll einem Test entsprechen, der objektiv ist und einem gehörten Gespräch entspricht. Das Verfahren in der Empfehlung P.862 entspricht einem solchen Test.

**MOS- LQE:** Dieser Wert soll von einem Netzwerkplanungsmodell berechnet werden. Dieses Modell muss das Ziel haben, die Qualität einer zu hörenden Applikation wiederzugeben.

**MOS- CQS:** Dieser Wert wurde in einem Labor erhoben. Das arithmetische Mittel wird aus gesammelten subjektiven Beurteilungen berechnet, welche zwischen 1 und 5 liegen, wie in der Empfehlung Rec. P.800 beschrieben. Subjektive Konversationstests wie in der Empfehlung P.800 liefern MOS-CQS.

**MOS- CQO:** Dieser Wert soll verwendet werden, wenn das berechnende Modell das Ziel hat, eine objektive Qualitätsbeur-

teilung eines Gesprächs zu ermitteln. Ein Beispiel dafür ist die Empfehlung Rec. P.562.

**MOS- CQE:** Bei einem Netzwerkplanungsmodell, welches der Qualität für Applikationen zur Konversation berechnet, wird MOS- CQE benutzt. Als Beispiel ist hier das Modell in der Empfehlung Rec. G.107 nach der Umrechnung zu nennen.

Der Zusammenhang von MOS- LQS, MOS- LQO und MOS- LQE ergibt sich wie folgt:

- ▶ Wird das zu testende Signal von einem Menschen subjektiv beurteilt, entspricht dies dem MOS- LQS.
- ▶ Wenn das gleiche Signal einem objektiven Algorithmus zur Verfügung gestellt wird, berechnet dieser einen MOS- LQO.
- ▶ Aus dem System, über das das Signal gesendet wird, können ermittelte Parameter an einen Algorithmus gegeben werden. Dieser hat eine Datenbank von differierenden Faktoren und kann daraus einen MOS- LQE liefern.

Die ITU-Empfehlung P.830 verfeinert diese Bewertungsmethode. Wichtigste Qualitätskriterien für die Übermittlung von Sprachinformationen sind die Verzögerungszeiten, Bitfehler-raten, Echos und Jitter. Da das Ohr auf Klangschwankungen und Sprachunterbrechungen sensibel reagiert, sollten die Verzögerungszeiten annähernd konstant sein. Die Sprachqualität wird durch die Verzögerung während der Übertragung nicht beeinträchtigt, es verschlechtert sich lediglich die bidirektionale Gesprächsqualität. Zu den wichtigen Qualitätskriterien zählen vor allem auch Unterbrechungen, d.h. Übertragungsverluste und Verlustraten. Besonders störend sind Aussetzer. Knacker alleine sind unangenehm, doch die merklichen Infor-



mationsverluste sind unerträglich. Leider kennt VoIP keine Fehlerkorrektur der Form „Bitte nochmal, habe nicht verstanden.“

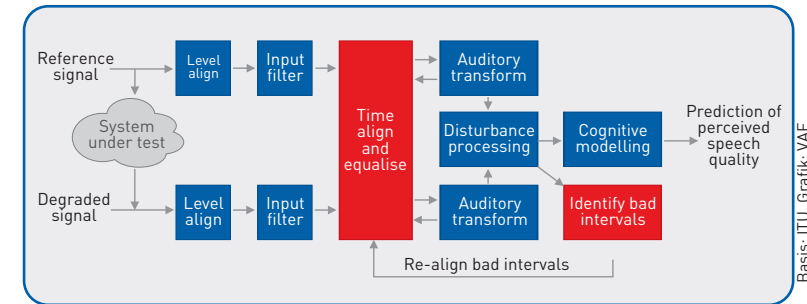
Echos entstehen in analogen Systemen am Übergang von Vierdraht- auf Zweidrahttechnik und irritieren den Sprecher durch die Sprachreflexion, worunter die Verständlichkeit leidet und vom Hörer höhere Konzentration gefordert wird.

R-Faktor	MOS-Wert	Subjektives Empfinden
90	4,34	Sehr zufrieden
80	4,03	Zufrieden
70	3,60	Wenige sind unzufrieden
60	3,10	Viele sind unzufrieden
50	2,58	Fast alle sind unzufrieden

Bewertung der Sprachqualität nach MOS und PESQ

### Objektive Bewertung der Sprachqualität nach PESQ

»Perceptual Evaluation of Speech Quality«, kurz PESQ, ist eine Methode zur objektiven Bewertung der Sprachqualität in der Telefonie im Frequenzbereich von 0,3 bis 3400 Hz. PESQ ist in der ITU-Empfehlung P.862 beschrieben und basiert auf den realen Bedingungen für eine Ende-zu-Ende-Sprachkommunikation.



Modell des PESQ-Algorithmus

Das Verfahren berücksichtigt unter anderem Paketverluste, Rauschen und den verwendeten Audio-Codec. PESQ liefert eine Bewertung der Sprachqualität im Bereich von »-0,5« bis »4,5«. Werte nahe »-0,5« bedeuten eine sehr schlechte Sprachqualität, während Werte nahe bei »4,5« eine sehr gute Sprachqualität signalisieren.

In den meisten Fällen liegen die Ausgabewerte zwischen »1« und »4,5«. Dies ist auf den ersten Blick überraschend, da der ITU-Skalenbereich für MOS bis »5« geht. Die Erklärung dafür ist aber einfach. PESQ simuliert einen Hörtest und wurde dafür entwickelt, dass durchschnittliche Ergebnis aller subjektiven Höreindrücke zu ermitteln. Statistiken beweisen, dass das am besten durchschnittliche Ergebnis, das man im Allgemeinen von einem Hörtest erwarten kann, nicht »5« ist. Es ist stattdessen rund »4,5«. Es scheint, dass die Testpersonen recht vorsichtig sind, für eine Probe eine »5« zu geben, selbst wenn es keine Verschlechterung gibt.

Wie die links stehende Abbildung zeigt, werden bei der PESQ-Analyse ein Referenzsignal und das geminderte Signal in das



System eingegeben. Im PESQ-System finden umfangreiche Vergleiche und Berechnungen statt, weshalb die Berechnung der Ergebnisse einige Zeit in Anspruch nehmen kann. Als Ergebnis erhält man neben dem oben beschriebenen PESQ-Wert eine Vielzahl von weiteren Parametern, die ausführlich die Sprachqualität beschreiben.

### MOS-Wert Subjektive Einstufung

5	Exzellente Qualität. Ohne jede Beanstandung.
4	Gute Qualität. Hohe Akzeptanz, erfüllt die allgemeinen Qualitätsansprüche der Benutzer.
3	Befriedigende Qualität. So wird sie vom Benutzer gerade noch akzeptiert. Allerdings strengt eine Kommunikation mit dieser Qualität die Benutzer bereits über Gebühr an.
2	Ungenügende Qualität. Die Kommunikation wird derart anstrengend, dass Benutzer sie nur dann hinnehmen (NICHT AKZEPTIEREN), wenn die Kommunikation auf keinem anderen Wege möglich ist.
1	Inakzeptable Qualität. Verbindungen kommen nicht zustande, Kommunikationsinhalte gehen gänzlich verloren.

MOS-Werte

### MOS-Wert von bis Subjektive Einstufung durch die Benutzer

MOS-Wert von	bis	Subjektive Einstufung durch die Benutzer
4,34	5	Alle Benutzer sind sehr zufrieden.
4,03	4,34	Nicht mehr jeder Benutzer ist sehr zufrieden, die meisten sind es.
3,60	4,03	Einige Benutzer sind unzufrieden.
3,10	3,60	Die Mehrheit der Benutzer ist unzufrieden.
2,58	3,10	Alle Benutzer sind unzufrieden.

Wertebereiche MOS

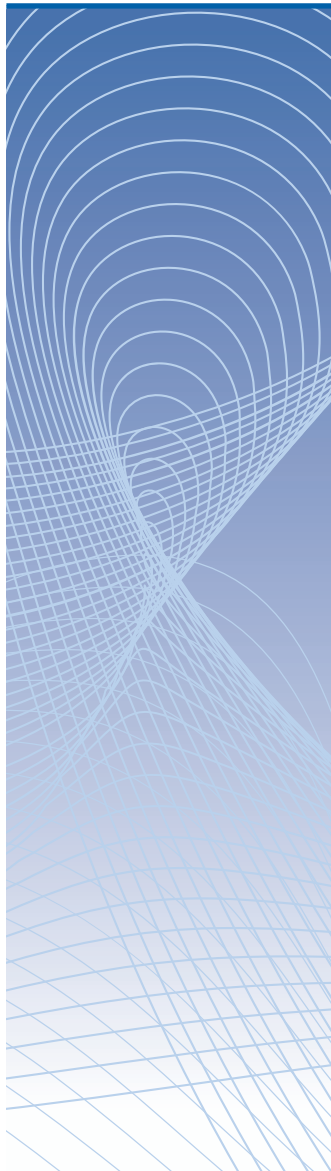
Codecbezeichnung	Kodierung	Bitrate (kBit/s)	Erreichbare MOS-Werte
G.711	PCM	64	bis 4,5
G.722.2 mit VAD und CNG	ADPCM	16, 32, 64	4,0 bis 4,5
G.721(1988), G.726, G.727	ADPCM	32	<3,5
G.723.1	ACELP	5,3	bis 3,8
G.723.1	MP-MLQ	6,3	bis 3,8
G.726, G.727	ADPCM	16, 24, 40	<3,5
G.728	LD-CELP	16, 12,8	<3,8
G.729	CS-ACELP	8	bis 4,3
G.729-A + VAD	CS-ACELP	8	bis 4,1
IS-54	VSELP	8	<3,8
IS-641	ACELP	7,4	<3,8
IS-96a	QCELP	8	<3,8
IS-127	RCELP	8	<3,8
Japanisches PDC	VSELP	6,7	<3,8

Codecs und erreichbare MOS-Werte

Wie die nachfolgende Tabelle zeigt, wirken sich die Paketverluste direkt auf die zu erzielende MOS-Werte aus.

Prozentualer Paketverlust	MOS-Wert von	bis
1%	4,0	4,2
2%	3,3	4,0
3%	3,1	3,3
4%	2,8	3,1
5%	2,6	2,8

Zusammenhang zwischen Paketverlusten und MOS-Werten



### Objektive Bewertung der Sprachqualität nach dem E-Model

Die ITU-T-Empfehlung G.107 beschreibt mit dem E-Model ein Berechnungsmodell zur Bestimmung von objektiven Qualitätsparametern für Sprachverbindungen. Hierbei wird der »Transmission Rating Factor« »R« unter Berücksichtigung folgender Einflussgrößen berechnet:

- ▶ Signal-Rauschabstand,
- ▶ simultane Beeinträchtigungen, beispielsweise Quantisierung oder Sidetone,
- ▶ Einflüsse der Verzögerung, beispielsweise absolute Verzögerung oder Echo,
- ▶ Beeinträchtigungen durch die Geräte, beispielsweise Codecs oder Paketverluste sowie
- ▶ Advantage-Factor, beispielsweise Mobilkommunikation oder Satellitenkommunikation.

Aus diesen Werten wird eine Vorhersage der Sprachqualität im Bereich von 0 bis 100 getroffen, die auf die MOS-Skala abbildbar ist. Das E-Model wurde ursprünglich als Planungsparameter für Telefonnetzbetreiber entwickelt. Inzwischen hat es sich als Quasi-Standard zur objektiven Beurteilung der Sprachqualität – im Gegensatz zur subjektiven Methode des MOS – durchgesetzt. Da sich der R-Faktor direkt auf den aus den Tests generierten Messwerten abbilden lässt, entspricht dieser Wert den realen Verkehrsparametern. Trotzdem ist eine Korrelation mit den MOS-Werten möglich.

Der beste zu erreichende theoretische R-Faktor beträgt 100. Dieser Wert berücksichtigt jedoch nicht die genutzten Codecs. Benutzt man beispielsweise einen typischen G.711-Codec in einem Standardraum, kann ein maximaler R-Faktor von unge-

fähr 93,2 erreicht werden. Die folgenden Ursachen tragen zu einer Verschlechterung des R-Faktors bei:

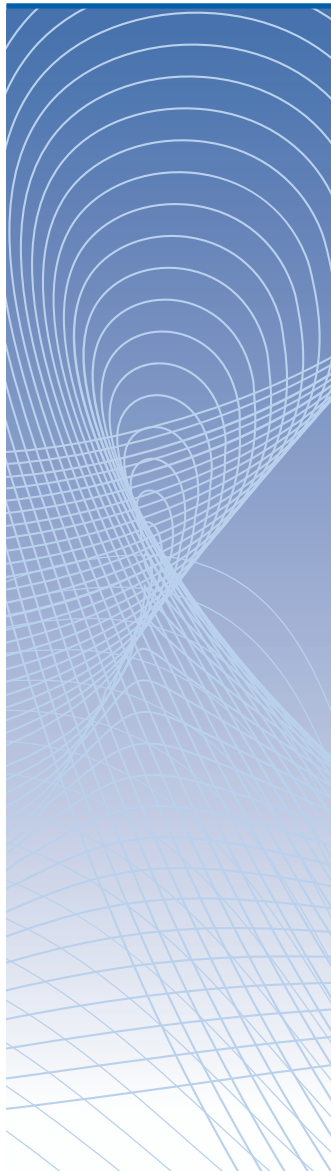
- ▶ Codec-Typ, da ein Codec mit einer höheren Kompression normalerweise einen schlechteren R-Faktor aufweist,
- ▶ Einschränkung der Übertragungsbandbreite durch das gesamte Übertragungssystem,
- ▶ Verzögerungen und Jitter im Netzwerk und Endgerät (mobiles WLAN-Telefon) und
- ▶ Paketverluste.

R-Faktor	MOS	Qualität
100	5,0	ausgezeichnet: Es ist keine Anstrengung nötig, um die Sprache zu verstehen.
80	4,0	gut: Durch aufmerksames Hören kann die Sprache ohne Anstrengung wahrgenommen werden.
60	3,1	ordentlich: Die Sprache kann mit leichter Anstrengung wahrgenommen werden.
50	2,6	mäßig: Es bedarf hoher Konzentration und Anstrengung, um die übermittelte Sprache zu verstehen
bis 0	bis 1,0	mangelhaft: Trotz großer Anstrengung kann man sich nicht verständigen.

Vergleich der Sprachqualitätsbewertung nach R-Faktor und MOS

### 1.2 Verzögerung

Der Begriff Verzögerung oder Delay, gemessen in Millisekunden, wird auch als Latenzzeit oder Latenz bezeichnet. Die Verzögerung ist das Zeitintervall zwischen dem Auftreten eines



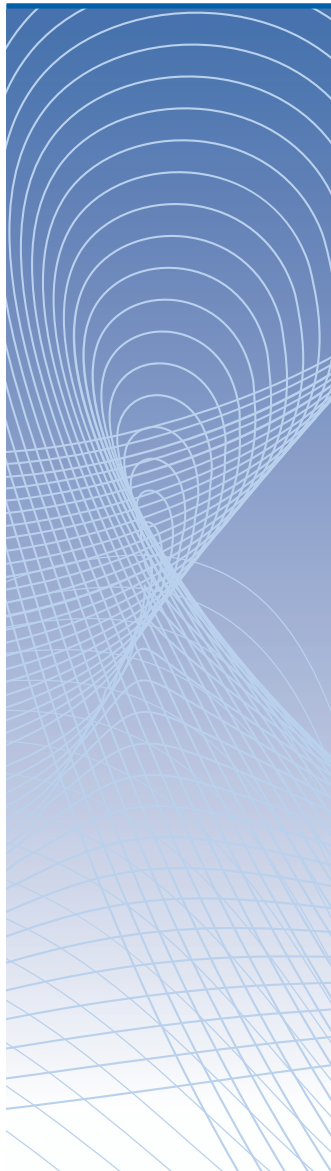
Ereignisses und dem Auftreten eines erwarteten Folgeereignisses, um das dieses verzögert wird. Im konkreten Fall ist die Verzögerungszeit der Zeitraum zwischen dem Sprechen und dem entfernten Hören der gesprochenen Nachricht. In Netzwerken wird die Verzögerung oft mit dem Begriff Round-Trip-Time (RTT) beschrieben. Der Round-Trip-Delay beschreibt die Gesamtverzögerung, also Hin- und Rückweg zwischen zwei IP-Endpunkten. Bei Voice-over-IP-Anwendungen ist der so genannte One-Way-Delay, also die Verzögerung in einer Richtung von Endpunkt zu Endpunkt, von Bedeutung. Verzögerung kann auf verschiedene Arten entstehen.

► **Verarbeitungsverzögerung (Processing-Delay):** Die Verarbeitungsverzögerung ist durch die zur Weiterverarbeitung des Signals benötigte Zeit bestimmt. Diese steht in Beziehung zur angewandten Signalverarbeitung und hängt von der verfügbaren Prozessorleistung ab, ist aber beim Echtzeitbetrieb auf die Dauer eines Frames beschränkt. Diese Verzögerung enthält folgende Zeitintervalle:

- Die notwendige Zeit, die zum Empfang eines Datenpakets inklusive des Sprach-Samples notwendig ist.
- Den Zeitraum zur Authentifizierung und Entschlüsselung der Pakete bei sicheren Übermittlungen.
- Den Zeitraum zum Durchlauf des Jitter-Puffers.
- Die Zeit, die der Sprachdecoder zur Verarbeitung der Signale benötigt. Diese Verzögerungsursache hängt vom verwendeten Sprach-Codec ab und tritt auf Grund des Frameing für die Blockverarbeitung und der Vorgriffe für die Parallelübertragung auf.
- Eine algorithmische Verzögerung entsteht durch Eigenschaften des zur Verarbeitung oder zur Übertragung des

Signals verwendeten Algorithmus. In die algorithmische Verzögerung fließen auch Verzögerungen ein, die durch die Vor- und Nachverarbeitung, die Echosperrung, die Rauschunterdrückung und die Filterung bedingt sind.

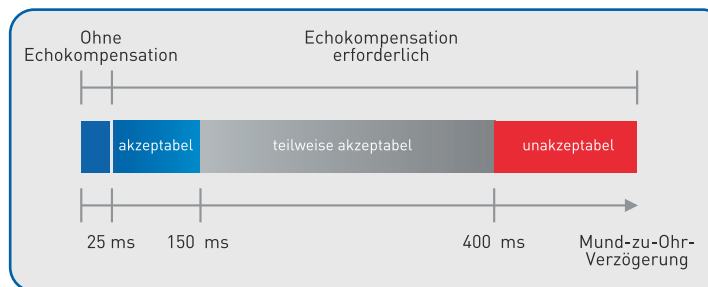
- **Serialisierungsverzögerung (Serialization-Delay):** Die Serialisierungsverzögerung ist die Zeit, die notwendig ist, um eine Dateneinheit über das betreffende Interface komplett auf das Medium zu schicken. Es gilt: Die Serialisierungsverzögerung ist das Produkt aus Datenmenge geteilt durch die Datenrate.
- **Netzwerkverzögerung (Network-Delay):** Die Netzwerkverzögerungen werden durch die physikalische Verzögerung der Übertragungsleitungen sowie der Queuing- und Pufferungsmechanismen in den Koppelkomponenten (Router, Switches oder Gateways) verursacht und variieren in ihrem Ausmaß. Die so genannte Durchlaufzeit setzt sich aus zwei Größen zusammen: einer konstanten oder nur leicht variierenden Netzverzögerung und schnellen Schwankungen der Verzögerung, gemeinhin als Jitter bekannt. Zur netzimmanenten Verzögerung gehört auch das Senden mehrerer Frames in einem Paket, eine Technik zur Reduzierung des Protokoll-Overheads bei IP-Übertragungen. Im Internet betragen die Netzverzögerungen in der Regel etwa 40 Millisekunden. Durch den Einsatz moderner Router und Switches mit höheren aggregierten Bandbreiten bzw. erheblich verbesserten Speicherfunktionen und steigenden Prozessorleistungen sinkt die netzbedingte Verzögerung kontinuierlich.
- **Ausbreitungsverzögerung (Propagation-Delay):** Die Ausbreitungsverzögerung entsteht durch die Signallaufzeiten in den jeweiligen Übertragungsmedien. Durch die Auswahl der richtigen Übertragungsmedien und -strecken kann dieser Parameter in Grenzen beeinflusst werden. Es gilt: Ausbreitungsverzö-



Netzwerk-Hardware	Verursache Verzögerung	Bemerkung
Netzwerkkarte	0,2 – 1 ms	Abhängig vom Netzwerk-Karten-Typ (10 MBit/s – 1000 MBit/s)
Klassischer Router	1 bis 10 ms	Abhängig vom Router-Typ
Switch (Layer 2)	0,2 – 1 ms	Abhängig vom Switch-Typ
Switch Layer 3)	0,2 – 1 ms	Hardware-basierter Switch
Switch Layer 3)	5 – 10 ms	Software-basierter Switch
Firewall oder Proxy Server	10 – 1000 ms	

Typische Verzögerungszeiten der verschiedenen Koppelkomponenten

gerung ist die Strecke im Medium dividiert durch die Ausbreitungsgeschwindigkeit. Für die Ausbreitungsgeschwindigkeit können als Faustregel fünf Mikrosekunden pro Kilometer angesetzt werden.



Basis: ITU, Grafik: VAF

Die typischen Verzögerungszeiten von Netzwerkkomponenten hängen stark von der eingesetzten Hard- und Software-Komponente ab.

Die Laufzeit setzt sich aus der Summe aller Verzögerungen, die während der Übertragung auftreten, zusammen. Jedes Netzelement erzeugt durch die internen Verarbeitungsprozesse eine bestimmte Verzögerungszeit. Für die Übertragung von Informationen in Echtzeit muss die Verzögerungszeit gering sein. Die Delay-Spezifikation gemäß G.114 der ITU-T definiert folgende Bereiche:

Je nach Nutzung variiert der Einfluss der Latenzzeit auf die Gesprächsqualität erheblich. Beispielsweise werden lange Verzögerungen bei einem Telefonat mit dem Mobiltelefon nicht so störend empfunden wie bei einem normalen schnurlosen Telefon, weil beim Mobiltelefon der Mobilitätsvorteil überwiegt.

Bei extremen Verzögerungen von über 200 Millisekunden können Anwender nur schwer telefonieren, ohne dass sie sich gegenseitig zwangsläufig unterbrechen. Die Standardisierungs-Abteilung der International Telecommunication Union, kurz ITU-T, empfiehlt für den G.114-Standard eine Verzögerung von unter 150 Millisekunden, um eine akzeptable Gesprächsqualität zu gewährleisten.

Bereich	Bedeutung
0 - 150 ms	akzeptabel für die meisten Echtzeitanwendungen
150 - 400 ms	teilweise akzeptabel unter Berücksichtigung der Qualitätsverluste
> 400 ms	unakzeptabel

Basis: ITU



## 1.3 Paketverlustrate

Die Paketverlustrate, auch Packet-Loss-Rate oder PLR genannt, ist in der Nachrichtentechnik ein Maß für die Übertragungsqualität einer elektronischen Datenverbindung. Die Paketverlustrate, die meist in Prozent angegeben ist, gibt an, wie viele Pakete eines Datenstroms zwischen einem Sender und einem oder mehreren Empfängern während der Übertragung verloren gegangen sind und somit nicht empfangen werden konnten. Die Paketverlustrate berechnet sich aus dem Verhältnis der Anzahl verloren gegangener zur Anzahl gesendeter Datenpakete:

$$\text{PLR} = \frac{\text{Anzahl der verlorenen Pakete}}{\text{Anzahl der gesendeten Pakete}}$$

Für die prozentuale Angabe muss nur der Quotient mit 100 multipliziert werden:

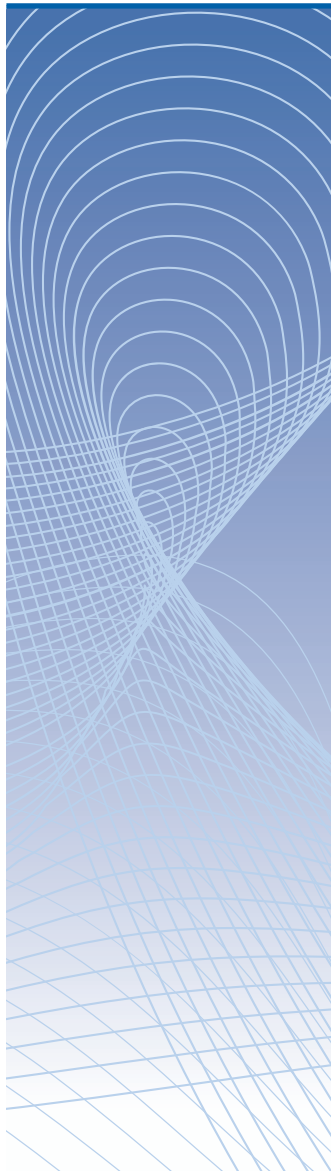
$$\text{PLR} = \frac{\text{Anzahl der verlorenen Pakete}}{\text{Anzahl der gesendeten Pakete}} \times 100$$

Um eine gute Verbindung zu haben, sollte dieser Fehlerwert so klein wie möglich sein. Optimal ausgelegte und gut administrierte IP-Backbones weisen heute in der Regel eine Paketverlustrate von unter 0,5 Prozent auf. (Gemäß ITU G.114 – Spezifikation ist eine Paketverlustrate von bis zu 5 Prozent noch akzeptabel.) Für viele Experten gilt jedoch eine Paketverlust-

trate von 5 Prozent für die Praxis als nicht mehr akzeptabel. Die Paketverlustrate ist maßgeblich abhängig von der Paketlänge und natürlich dem darin transportierten Inhalt. Verliert das System große Pakete mit komprimiertem Inhalt wirkt sich das deutlicher aus als kleine Pakete mit wenig Kompression. Somit ist die Paketgröße ein entscheidender Einflussfaktor.

Die folgenden Aspekte sind jedoch bei der Betrachtung der Paketverlustrate zu beachten, denn diese beeinflussen unter Umständen die Sprachqualität.

- ▶ Es können zwei unterschiedliche Arten der Paketverluste unterschieden werden:
  - ▶ Der Paketverlust einzelner Datenpakete über einen bestimmten Zeitraum,
  - ▶ Der Paketverlust mehrerer aufeinander folgender Datenpakete über einen bestimmten Zeitraum (Consecutive-Packet-Loss).
  - ▶ Der Verlust von mehreren, aufeinander folgenden Datenpaketen (Daten-Burst) ist vom Empfänger nicht mehr kompensierbar und wirkt sich als Übertragungsstörung aus. Vereinzelt Paketverluste werden vom Gehör/Gehirn interpoliert und fallen dem Zuhörer nicht auf.
  - ▶ Paketverluste wirken sich umso stärker aus, je länger die so genannte Payload, also der Sprachanteil im Paket, ist. Verloren gegangene längere Sprachanteile nimmt der Zuhörer wahr. Der Verlust kürzerer Sprachanteile wird vom Gehirn des Zuhörers interpoliert und somit subjektiv nicht bemerkt.
  - ▶ Codecs weisen eine gewisse Toleranz gegenüber Paketverlusten auf. In Abhängigkeit vom genutzten Codec der Sprachanwendung bemerkt der Zuhörer die aufgetretenen Paketverluste nicht.



▶ Die Auswirkungen von Paketverlusten sind bei der Übermittlung von Tönen wesentlich höher als bei der Übertragung von Sprache. Das menschliche Ohr erkennt Daten- beziehungsweise Informationsverluste bei der Übertragung von Sprache wesentlich schlechter als bei der Übermittlung von Tönen. Aus diesem Grund sind Paketverluste bei Tonübertragungen deutlich störender als bei der Sprachübermittlung und erklären auch, warum es immer wieder bei der Fax- und Modemübermittlungen über IP zu Problemen kommt. Bei beiden Verfahren verläuft die Kommunikation über Töne.

Beim Messen von Paketverlusten sollten folgende Aspekte beachtet werden:

- ▶ Verhältnisermittlung zwischen gesendeten und empfangenen Paketen,
- ▶ Verteilung der verlorenen Pakete auf der Zeitachse sowie
- ▶ Festlegung der Kriterien, die ein Paket als verloren bezeichnen.

Die Verhältnisermittlung zwischen gesendeten und empfangenen Paketen ist einfach zu berechnen. Der Sender integriert in jedem abgehenden Paket eine Sequenznummer. Anhand der Lücken in den Sequenznummern ermittelt der Empfänger die fehlenden Pakete.  $S_{max}$  entspricht der höchsten und  $S_{min}$  der kleinsten in einem Zeitintervall beobachteten Sequenznummer. Der Wert  $n$  entspricht der Anzahl zwischen  $S_{min}$  und  $S_{max}$  liegender Sequenznummern. Aus diesen Werten errechnet sich die Verlustrate  $r$  wie folgt:

$$r = \frac{n}{S_{max} - S_{min}}$$

Schwieriger als das Berechnen der Verlustrate ist die Festlegung von Entscheidungskriterien, die den Verlust von Paketen eindeutig definieren. Die Antwort hängt immer von der Art der Anwendung ab. Bei nicht zeitkritischen Anwendungen kann lange auf ein verspätetes Paket gewartet werden und Reihenfolgevertauschungen haben einen geringen Einfluss. In solchen Anwendungen wird ein Paket erst als verloren betrachtet, nachdem ein gewisses Zeitintervall überschritten ist. VoIP-Anwendungen hingegen reagieren sehr empfindlich auf Verzögerung und Jitter. Aus diesem Grund werden verspätete Pakete meist als verloren gewertet. Ähnlich verhält es sich mit vertauschten Paketen: Reihenfolgevertauschungen können höchstens innerhalb des Synchronisationspuffers korrigiert werden. Ist dieser nicht groß genug, wird auch ein solches Paket als verloren betrachtet.

In der Default-Einstellung wird bei den meisten VoIP-Anwendungen der Codec G.711 genutzt. Um hörbare Störungen zu vermeiden erfordert dieser Codec eine Paketverlustrate von weniger als einem Prozent. Die Paketlänge ist jedoch entscheidend für die Paketverlustrate. Hier gilt: Je kürzer die Pakete sind, umso besser ist dies für die Sprachqualität.

## 1.4 Jitter

Als Jitter, gemessen in Millisekunden, bezeichnet man allgemein ein Taktzittern bei der Übertragung von Digitalsignalen beziehungsweise eine leichte Genauigkeitsschwankung im Übertragungstakt. Der Jitter macht sich durch ein Echo im Telefonat bemerkbar. In der Netzwerktechnik wird mit Jitter au-

ßerdem die Varianz der Laufzeit von Datenpaketen bezeichnet.

Der Jitter-Effekt ist insbesondere bei Multimedia-Anwendungen störend, da dadurch Pakete zu spät eintreffen können, um noch zeitgerecht mit ausgegeben werden zu können. Dies wirkt sich wie eine erhöhte Paketverlustrate aus. Treffen die Pakete einer Audioverbindung regelmäßig beim Empfänger ein, können diese direkt in Audiosignale umgesetzt werden. Da die Verzögerungen bei der Übertragung nicht konstant sind, entstehen

#### ▶ BANDBREITE AUS SICHT EINES NUTZERS

Betrachtet man das Thema »Bandbreite« aus der Sicht eines Nutzers, so sind typischerweise die folgenden Erwartungen zu erfüllen.

- ▶ **Hohe Verfügbarkeit:** Der Nutzer erwartet, dass sein Dienst immer funktioniert. Diese Erwartung ist in der Realität nicht zu erfüllen, jedoch kann man ihr durch redundante, hochverfügbare Systeme sehr nahe kommen.
- ▶ **Kurze Wiederherstellungszeit:** Ist ein Dienst nicht verfügbar, so erwartet der Nutzer seine Wiederherstellung innerhalb einer von ihm als akzeptabel empfundenen Zeit. Je nach Dienst gibt es eine gewisse Toleranz des Nutzers, die von seinem Charakter, seiner Einsicht und seiner Bewertung der Komplexität des Dienstes geprägt wird. So wird typischerweise ein Ausfall des Telefons für fünf Minuten durchaus höher bewertet als ein Ausfall eines Web-Servers für fünf Minuten.
- ▶ **Adäquate Bedienzeit:** Der Nutzer möchte von seinem Dienst so schnell bedient werden, dass sein Handlungsablauf nicht behindert wird.
- ▶ **Hoher Durchsatz:** Der Nutzer möchte adäquate Bedienzeiten erfahren, auch wenn der Dienst stark nachgefragt wird.

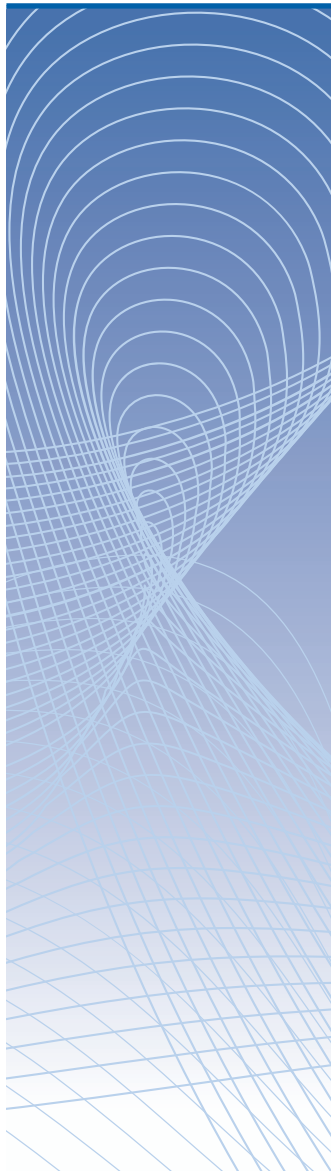
Lücken im abgespielten Signal. Die Differenz zwischen den Verzögerungen einzelner Pakete wird als Jitter oder Verzögerungsschwankung bezeichnet.

Zur Vermeidung von Lücken im Signal müssen die empfangenen Daten in einem Zwischenspeicher abgelegt werden. Dieser Zwischenspeicher hat die Aufgabe, die Lücken zwischen verspäteten Paketen zu kompensieren. Die Größe dieses Zwischenspeichers, also des Synchronisationspuffer oder Jitter-Buffer, kann ein oder mehrere Kodierungsframes umfassen. Jitterbuffer sorgen für eine Verbesserung des Jitter-Verhaltens, vergrößern jedoch die Gesamtverzögerung. Zur Verbesserung der Sprachqualität werden oft dynamische Jitter-Buffer eingesetzt, die Länge der Buffer orientiert sich am gemessenen Jitter.

Einfacher als die Verzögerung sind die Schwankungen zwischen den Verzögerungen einzelner Pakete, also der Jitter, zu messen. Für zwei aufeinander folgende Pakete p und q mit Sendezeiten  $p_s$  und  $q_s$  und Empfangszeiten  $p_e$  und  $q_e$  lässt sich der Jitter wie folgt berechnen:

$$\text{Jitter} = (q_e - p_e) - (p_s - q_s)$$

Sprachdecoder haben die Aufgabe, den Datenfluss aufrecht zu halten. Treffen Datenpakete mit zu kleinem Jitter ein, muss der Decoder dies durch das Einfügen von zusätzlichen Bits ausgleichen. Ist der Jitter aber groß, sind größere Jitter-Buffer erforderlich. Je größer der Jitter-Buffer ist, desto größer wird aber auch die Verzögerung. Andererseits führt ein zu klein gewählter Jitter-Buffer schnell zu Paketverlusten.



Der Jitter lässt sich durch intelligente Buffer kompensieren. Diese gehen jedoch wieder zu Lasten der Laufzeit. Leider schließen sich die QoS-Parameter an vielen Stellen gegenseitig aus, was das Thema in der Praxis so komplex macht.

### 1.5 Datendurchsatz

Der Datendurchsatz oder die Performance hat die Einheit Bit/s (kBit/s, MBit/s, GBit/s) und zeigt die Ende-zu-Ende-Transportrate an.

Wenn die Performance einer Netzwerkstrecke unterhalb des erforderlichen Datendurchsatzes liegt, bedeutet dies immer ein Ansteigen der Paketverlustrate und damit eine Verringerung

der Güte. Jede Applikation benötigt eine definierte Bandbreite, wenn diese Informationen in Echtzeit übertragen werden soll. Typische Werte für benötigte Bandbreiten sind:

Netzwerk	Geschwindigkeit
PSTN (herkömmliche Sprachnetze)	56 kBit/s / 64 kBit/s
Videokonferenzen (niedrige Qualität)	128 - 384 kBit/s
Videostreaming (mittlere bis hohe Qualität)	384 kBit/s - 6 MBit/s

Im direkten Kontext zu den benötigten Bandbreiten stehen die verwendeten Codecs und Kompressionsalgorithmen:

Standard	Algorithmus	Bitrate (kBit/s)	Delay (ms) ohne Kanal-Delay	Sprachqualität (MOS)
G.711	PCM	64	<<1	ausgezeichnet
G.723.1	MP-MLQ/ACELP	5,3, 6,3	67 - 97	gut bis ordentlich
H.728	LD-CELP	16	<<2	gut
G.729	CS-ACELP	8	25 - 35	gut
G.729 Annex A	CS-ACELP	8	25 - 35	gut
G.722	Sub-Band ADPCM	48, 56, 64	<<2	gut
G.726	ADPCM	16, 24, 32, 40	60	gut bis ordentlich
G.727	AEDPCM	16, 24, 32, 40	60	gut bis ordentlich

Voice-Encoding nach der ITU-T-Spezifikation

## 2. Sprachqualität in IP-Netzen

IP-Netze sind von Hause aus nur sehr bedingt für die Sprachübertragung geeignet. Mit Techniken wie dem Bandbreitenmanagement werden sie für die Anforderungen von Real-Time-Anwendungen fit gemacht.

Im ursprünglichen Design von TCP/IP spielte ein Bandbreitenmanagement eine untergeordnete Rolle. TCP/IP war für den Transport von Daten für Anwendungen wie E-Mail oder Internet gedacht und nicht für Sprache oder Video. Aus diesem Grund besteht bei den meisten auf dem IP-Protokoll basierenden Netzwerken immer eine Unbestimmtheit und Unvorhersehbarkeit der Datenübertragungsparameter Laufzeit,

Verzögerungen, Jitter und Datenverlust. Die meisten »datenorientierten« Anwendungsprogramme tolerieren solche Schwankungen und stellen keine hohen Ansprüche. In aller Regel steht genügend Zeit zur Verfügung, eventuelle Fehler durch Rekonstruktion beziehungsweise durch ein wiederholtes Anfordern der betroffenen Datenpakete auszugleichen.

Netzwerke auf Basis des Internet-Protokolls gehören zu den sogenannten Best-Effort-Netzwerken. Das Netzwerk unterstützt viele Anwender und Anwendungen und es soll die Daten ordnungsgemäß zu übertragen. Dies geschieht aber von Hause aus ohne eine bestimmte Dienstgüte zu garantieren. Für Applikationen wie File-Transfer (ftp) oder E-Mail reichen die Übermittlungscharakteristiken von Best-Effort-Netzwerken völlig aus. Der Vorteil dieser Netze besteht darin, dass die Komple-

xität auf die Endgeräte reduziert wird und das eigentliche Transportnetz relativ einfach und preiswert realisiert werden kann. Dieser Lösungsansatz bietet darüber hinaus erhebliche Skalierungsvorteile.

Die Vorteile der Skalierbarkeit und der Einfachheit des Netzkonzepts wirken sich jedoch bei Echtzeitanwendungen negativ aus. Erfordert ein Dienst eine Bandbreite, die die momentan zur Verfügung stehende Netzbandbreite übersteigt, weist das Netzwerk diesen Dienst oder Nutzer nicht ab. Stattdessen wird versucht, die Performance aller Nutzer zu reduzieren, um so die notwendige Bandbreite bereitstellen zu können. Dies geschieht durch das Verzögern der Übertragung beziehungsweise das Verwerfen einzelner Datenpakete.

Da die herkömmlichen Standardanwendungen der TCP/IP-Welt mit diesen variablen Bandbreiten und der sich ständig ändernden Netzperformance gut umgehen können, bemerkt diese der Anwender nicht. Die Datenpakete erreichen den Zielrechner innerhalb einer gewissen Zeit. Verzögerte Datenpakete werden an den Zielrechner ausgeliefert, da die meisten Anwendungen das verbindungsorientierte TCP-Transportprotokoll nutzen.

## ▶ KAPAZITÄTSBEDARF VON VOIP-VERBINDUNGEN

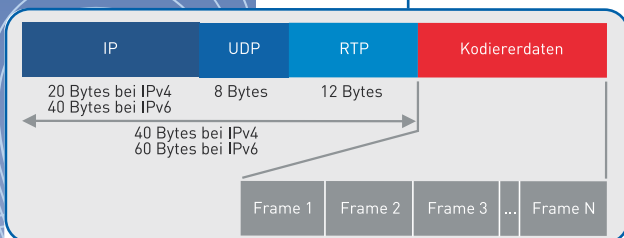
Bei Voice-over-IP-Anwendungen stellt sich immer wieder die Frage, welche Kapazitätsanforderung eine paketorientierte Sprachverbindung an ein Datennetz stellt. Das Session-Initiation-Protocol (SIP) nutzt zur Signalisierung ein http-ähnliches Protokoll auf Basis von TCP/IP. Die Nutzerinformationen (Sprachdaten) werden auf Basis des Reliable-Transport-Protocol (RTP) übertragen. Den Löwenanteil an Kapazität beziehungsweise Bandbreite benötigen natürlich die Sprachdaten, die entsprechend den verschiedenen Codierverfahren G.711, G.729 oder G.723.1 übermittelt werden. Beim SIP läuft der Verbindungsaufbau quasi zweistufig ab. Erst wird mit Hilfe von SIP

der Verbindungswunsch dem Proxy-Server signalisiert, um dann im Ergebnis die Mediaströme auf Basis von RTP direkt zwischen den Endgeräten auszutauschen. Um den Kapazitätsbedarf der Mediaströme nach erfolgreichem Verbindungsaufbau berechnen zu können, muss der aktuell eingesetzte Protokollstack bekannt sein, über den die Sprachdaten in pa-

ketorientierten Datennetzen übertragen werden. Im Normalfall besteht dieser aus dem Ethernet (Schicht 2) und dem darüber liegenden TCP/IP-Stack (Schichten 3 bis 7). Der Paket-Overhead berechnet sich im Standardszenario daher wie folgt:

Art	Länge
Payload (Nutz- beziehungsweise Sprachdaten)	variabel
RTP-Header	16 Byte Minimum
UDP-Header	8 Byte
IP-Header	20 Byte Minimum
Ethernet-Header, Trailer	18 Byte Minimum
Synchronisation	8 Byte

Der Payload ist abhängig von der gewählten Paketgröße. Diese kann in der Praxis variieren. Hinzu kommt, dass nicht jeder Codec jede Paketgröße unterstützt.



Overhead bei VoIP-Anwendungen

Codec	G.711 (64kBit/s)		G723.1 (6,3kBit/s)		G.729a (8kBit/s)	
	20 ms	30 ms	30 ms	20 ms	30 ms	40 ms
Paketlänge						
Reine IP-Bandbreite	80,0	74,6	17,0	24,0	18,6	16,0
IP-Bandbreite & RTP/UDP/IP-Kompression	95,2	84,8	27,2	39,2	28,8	23,6
Standard-PPP-Bandbreite	84,8	77,8	20,2	28,8	21,8	18,4
Standard-PPP-Bandbreite mit RTP/UDP/IP-Kompression	68,0	66,6	9,06	12,0	10,6	10,0
MPLS	81,6	75,7	18,2	25,6	19,7	16,8

Die Kenntnis des Protokollstacks wie auch der aktuellen Payload sind die Voraussetzung zur Berechnung des Kapazitätsbedarfs.

### Berechnung

In diesem Beispiel soll der Kapazitätsbedarf einer VoIP-Verbindung für eine G.711-Sprachcodierung mit 64 kBit/s netto berechnet werden. Dabei wird eine übliche Payload-Größe von 240 Byte angenommen. Die Kommunikation findet in unserem Berechnungsbeispiel in einem lokalen Ethernet-Netz statt. Aus diesen Vorgaben lässt sich folgendes berechnen:

▶ **Größe eines Sprachpakets:** (Ethernet-Header) + (IP/UDP/RTP-Header) + Payload = 302 Byte (2416 Bit)

▶ **Nettorate des Codecs** = 34 Pakete/s

Daraus lässt sich unmittelbar die erforderliche Kapazität der Verbindung für die Sprachdatenübertragung errechnen:

▶ **Erforderliche Kapazität der Verbindung:** (Größe des Sprachpakets = 2416 Bit) x (Nettorate des Codecs = 34 Pakete/s) = 82,144 kBit/s.

Darüber hinaus sorgt TCP für:

- ▶ Einen fehlerfreien Datentransfer,
- ▶ die Reihenfolgerichtigkeit der Daten,
- ▶ die bedarfsgesteuerte Übermittlung verloren gegangener Datensegmente,
- ▶ das Verwerfen von Datenduplikaten,
- ▶ eine Flusskontrolle und
- ▶ eine dynamische Anpassung an die verfügbare Bandbreite.

Echtzeitapplikationen wie VoIP sind nicht immun gegen Veränderungen der verfügbaren Bandbreite während des Datentransfers. Außerdem ist es bei diesen Anwendungen nicht notwendig, verloren gegangene Datenpakete erneut zu übermitteln. Grundsätzlich kann das SIP-Protokolle sowohl über die TCP- oder die UDP-Mechanismen auf der Transportebene übermittelt werden. Das RTP-Protokoll arbeitet immer auf Basis des UDP-Protokolls.

Das verbindungslose UDP-Protokoll ist dank geringerem Overhead wesentlich schneller als TCP und bietet für zeitsensitive Applikationen eine höhere Effizienz. Das UDP-Protokoll bietet keine Mechanismen für:

- ▶ eine Reihenfolgerichtigkeit,
- ▶ eine Empfangsgarantie und
- ▶ eine Flusskontrolle.

Aus diesem Grund erfordern Echtzeitapplikationen vom Netzwerk eine gewisse Konstanz an verfügbarer Bandbreite.

## 2.1 Network-Service-Quality

Die Dienstqualität ist kein Selbstzweck, sondern dient dem Ziel, die Erwartungen der Nutzer an die Dienste zu erfüllen. Diese

Erwartungen sind nicht statisch, sondern ändern sich mit der Zeit durch die von den Nutzern eingenommene Perspektive, mit welchem Niveau von Qualität er im Kontext seiner jeweiligen persönlichen Erfahrungen zufrieden ist. Der Begriff Network-Service-Quality, kurz NSQ, beschreibt einen Mechanismus, der dafür sorgt, dass die Performance eines Netzwerks nach bestimmten Regeln den jeweiligen Datenflüssen im Transportnetz zugeordnet wird.

Das Management der Network-Service-Quality wurde erstmals in den zellbasierten ATM-Netzen realisiert. Heute wird der Begriff hauptsächlich in Zusammenhang mit paketbasierten Netzen genutzt. NSQ stellt sich immer in den Bezug zu den Verkehrscharakteristiken traditioneller Netzwerke. Somit stehen die spezifischen Parameter effektive Bandbreite, Verzögerung, Paketverluste und Jitter die die IP-Telefonieapplikationen beeinflussen, im Mittelpunkt von NSQ.

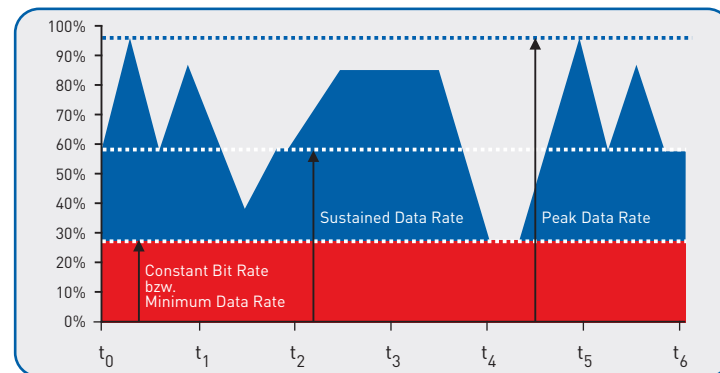
## 2.2 Effektive Bandbreite

Als effektive Bandbreite bezeichnet man die Bandbreite, die über den gesamten Netzwerkpfad Ende zu Ende zu einem bestimmten Zeitpunkt für die betreffende Anwendung beziehungsweise den jeweiligen Datenfluss zur Verfügung steht. In einem Netzwerk, beziehungsweise bei Netzwerkverbindungen, verändert sich die effektive Bandbreite permanent. Sie ist von der Anzahl gleichzeitiger Datenströme abhängig. In den meisten Fällen wird diese durch einige wenige überlastete Netzverbindungen oder Koppellemente bestimmt.

Bandbreiteneinschränkungen lassen sich am besten in Analogie zum Verkehrsfluss einer Autobahn erklären, der durch

Unfälle oder starkes Verkehrsaufkommen zu unterschiedlichen Tageszeiten nur eingeschränkt erfolgt. Dabei wird sicherlich ein konstantes Verkehrsaufkommen in Zeitintervallen stattfinden. Es wird zeitliche Verkehrsverdichtungen geben und es wird sich im zeitlichen Mittel eine bestimmte Auslastung des Transportmediums feststellen lassen.

Überträgt man die Beispiele aus dem Bild des Straßenverkehrs auf ein Datenübertragungsmedium, so sind die zugeordneten Messgrößen: Minimum- bzw. Constant-Bit-Rate (Mindestauslastung, die konstant über den gesamten Betrachtungszeitraum vorliegt), Peak-Data-Rate (Spitzenwerte der Auslastung) und Sustained-Data-Rate (Durchschnittliche Auslastung). Die meisten Sprachcodecs benötigen einen Durchsatz zwischen 35 kBit/s und 100 kBit/s pro Kommunikationsverbindung. Für den Einsatz von VoIP-Telefonen und Telefonieanwendungen sollte daher die effektive Bandbreite immer die aktuell notwendige Bandbreite übersteigen. Nur so steht den Nutzern immer ein konstanter Netzdurchsatz für seine Kommunikationsbeziehungen zur Verfügung. Aus diesem Grund



Bandbreitenmessgrößen

Grafik: VAF

müssen die Netzadministratoren dafür sorgen, dass in den Netzwerken die Bandbreite mit Hilfe von Managementfunktionen umfassend überwacht und kontrolliert wird.

## 2.3 Bandbreitenmanagement

Zunächst stellt sich die Frage, welche grundsätzlichen Möglichkeiten es in TCP/IP-Netzwerken gibt, ein effektiveres Bandbreitenmanagement durchzuführen. Um ein effektives Bandbreitenmanagement durchführen zu können, gibt es unterschiedliche methodische Vorgehensweisen. Grundsätzliche Analogien zu verschiedenen Problemen, die sich mit der Verteilung endlicher Ressourcen auseinandersetzen, zeigen, dass die Priorisierung von bestimmten Anforderungen ein möglicher Weg ist, begrenzte Ressourcen vernünftig zu verteilen. Dies setzt eine Art der Einordnung beziehungsweise Klassifizierung voraus. Ein weiterer Weg ist es, Ressourcen exklusiv zuzuordnen. Man kann Datenkanäle zwischen zwei Kommunikationsendpunkten reservieren, auf denen die Daten exklusiv übertragen werden. Dies wird im Allgemeinen dadurch erreicht, dass vorher signalisiert wird, dass diese Ressourcen exklusiv zur Verfügung stehen müssen.

Auch eine intelligente Verwaltung der Ressourcen mittels Einteilung und Zuteilung der Anfragen in unterschiedliche Abfertigungs- oder Warteschlangen erlaubt ein vernünftiges Zuteilen endlicher Ressourcen. Präventive Maßnahmen, bevor es zu Engpässen bestimmter Ressourcen kommt, beispielsweise eine intelligente Bandbreitenverwaltung, erlauben es, Engpässe von Ressourcen gar nicht erst aufkommen zu lassen, indem entsprechende Maßnahmen aktiv ergriffen werden.



Bestimmte Regeln (Policies), die auf die endlichen Ressourcen angewandt werden, erlauben eine effektive Zuteilung von knappen Ressourcen. Auch das aktive Reagieren zur Vermeidung von eventuellen Bandbreitenengpässen ist ein Weg, um Daten effektiv zu übertragen. Die Vermeidung von Bandbreitenengpässen (Congestion-Avoidance) beinhaltet die Überwachung der Auslastung des Netzwerks sowie das gezielte Verwerfen oder eine Abstufung der Priorität von Datenpaketen, um einem Bandbreitenengpass entgegenzuwirken.

Ein Bandbreitenmanagement für VoIP lässt sich mit folgenden Mitteln durchführen:

- ▶ Überdimensionierung der Netze,
- ▶ Rightsizing der Netze,
- ▶ Verkehrsklassen-Klassifizierung (Priorisierung),
- ▶ Signalisierungs-Mechanismen und -Verfahren,
- ▶ Warteschlangen und intelligente Warteschlangenverwaltung (Queuing und Scheduling),
- ▶ intelligente Bandbreitenverwaltung (Traffic-Policing),
- ▶ Bandbreitenzuweisung (Admission-Control) sowie
- ▶ Quality-of-Experience.

Die verschiedenen Techniken werden in den nachfolgenden Kapiteln dargestellt.

### 3. Überdimensionierung und Rightsizing

Die einfachste Methode, Bandbreitenengpässe zu vermeiden, ist Bandbreite im Überfluss zur Verfügung zu stellen und so dafür zu sorgen, dass sie aller Voraussicht nach nicht knapp werden kann.

Die gegenwärtigen und zukünftig zu erwartenden Kosten der Netzwerkkomponenten haben sich auch in den vergangenen Jahren stark zum Vorteil der Nutzer entwickelt. Hier wirkt sich indirekt auch das aus der Computertechnik bekannte Mooresche Gesetz – die Verdoppelung der Leistungsfähigkeit von Chips in 18 Monaten – vorteilhaft für die Netzwerkkomponenten aus. Ethernet-Netzwerke, die im Core-Bereich mit einem Mehrfachen an 10-Gigabit-Ethernet ausgerüstet sind, sind keine Seltenheit mehr. Der Preisverfall der Ethernet-Technik macht es möglich. Daher besteht heute in der Überdimensionierung des Netzes die einfachste Möglichkeit, für ausreichende Bandbreite zu sorgen.

Die Übertragungsgeschwindigkeiten in der Netzwerkperipherie – elektrische oder optische Leitungen wie auch Funkstrecken – und die Verarbeitungsgeschwindigkeit der angrenzenden Netzknoten wie Switches oder Router muss bei diesem Lösungsansatz stark überdimensioniert sein. Dies bietet den Vorteil, dass eine Verringerung der Wartezeiten in den Paket-Vermittlungsknoten erreicht wird. Grundlage einer Überdimensionierung der Netze ist die Homogenität der Netzkomponenten. Eine bewusste Beschränkung auf eine möglichst

einheitliche Ausstattung der Netzwerkkomponenten ist ein wichtiger Schlüssel, um eine leistungsfähige und qualitätsgesicherte Prozessorganisation zu erreichen.

Bei Überdimensionierung sind in den Teilnetzen keine zusätzlichen administrativen Maßnahmen nötig. Dies gewährleistet eine unverändert einfache Wartung und Fehlersuche und vermeidet Abhängigkeiten von herstellereigenen Sonderlösungen. Diese Methode ist immer anwendbar und erfordert keine Veränderung der bisher verwendeten Funktionsprinzipien, also auch der Übermittlungsprotokolle und Algorithmen in den Vermittlungsknoten.

In der Praxis ist die Überdimensionierung des Netzes nur kurzfristig sinnvoll. In Zukunft ist mit immer höher werdenden Nutzdatenraten zu rechnen. Moderne Applikationen erfordern immer höhere Bandbreiten und neue Anwendungen wie Videokonferenzen oder Videodienste fressen die verfügbaren Bandbreiten kontinuierlich auf. Die Situation ist ähnlich wie der Besitz einer neuen großen Festplatte: Sofort wird diese mit neuem Datenmaterial gefüllt, so dass sie in kurzer Zeit wieder voll ist, denn die Verfügbarkeit schafft neue Möglichkeiten, die dann auch genutzt werden.



Darüber hinaus sind selbst in einem lokalen Netzwerk die Zeitpunkte und Orte des Verkehrs nur schwer abschätzbar. Ein Zusammenbruch eines zentralen Knotens kann trotz integrierter Fehlertoleranz zu Ausfällen führen. Auch kommt es in Netzen, deren mittlere Bandbreitenausnutzung aller Links auf 50 Prozent dimensioniert ist, nach wie vor in gewissen Hochlastfällen zu Verstopfungen, die zu erhöhten Verzögerungen, Verzögerungsschwankungen und Paketverlusten führen.

Bei diesem Lösungsansatz ist unbedingt ein kontinuierliches Monitoring der verfügbaren Bandbreite erforderlich. Auf Basis der entsprechenden Kennzahlen wird bestimmt, ob die Bandbreitenansprüche erfüllt werden oder ob Handlungsbedarf besteht. Darüber hinaus ist ein laufendes Monitoring der Bandbreite hilfreich, um eine schleichende Degeneration der verfügbaren Netzressourcen frühzeitig zu erfassen und aktiv entsprechende Maßnahmen einleiten zu können.

#### ▶ ANMERKUNGEN: ÜBERDIMENSIONIERUNG

- Die Überdimensionierung erfordert nur rudimentäres Wissen über den Bandbreitenbedarf einzelner Anwendungen, jedoch müssen Fehlertoleranzmechanismen in das Netzwerk integriert werden, da sonst dennoch Engpässe und Dienstemängel durch unkalkulierbare Lastspitzen entstehen können.
- Für die vorsorgliche Bereitstellung von Überkapazitäten im LAN und insbesondere für überdimensionierte WAN-Strecken entstehen zusätzliche Betriebskosten.
- Unerwarteten Engpässen bei zunehmender Auslastung freier Ressourcen durch neue Anwendungen kann mittels Monitoring vorgebeugt werden.

Überdimensionierung ist in der Praxis nicht nur kostspielig sondern auch nur bedingt geeignet. Befinden sich auf einer mehrspurigen Autobahn und auf allen Fahrstreifen viele LKWs, behindern diese trotzdem den Verkehr der deutlich schnelleren PKWs. Es gilt somit: Bandbreite alleine ist keine Lösung!

### 3.1 Überdimensionierung im WAN

Von einigen Unternehmen werden sogenannte »schwarze Fasern« als Punkt-zu-Punkt-LWL-Verbindungen zwischen Unternehmensniederlassungen genutzt. Da die LWL-Fasern mit Hilfe der Wellenlängenmultiplextechnik, auch Wavelength-Division-Multiplex oder kurz WDM genannt, extrem hohe Durchsätze bereit stellt, kann diese Lösung in Einzelfällen genutzt werden. Generell werden die WAN-Verbindungen jedoch durch externe Serviceunternehmen realisiert. Da eine Überdimensionierung der WAN-Strecken nur auf Basis spezieller Service-Level-Agreements, kurz SLAs, realisiert werden, entstehen erhebliche Betriebskosten. Aus Kostengründen scheidet daher eine Überdimensionierung der WAN-Ressourcen als Lösungsansatz meist aus.

### 3.2 Rightsizing

Unter Rightsizing wird der Ansatz verstanden, bei vermeintlich knappen Ressourcen bevorzugt zu versuchen, die Knappheit zu beseitigen, anstatt eine Verwaltung der vermeintlichen Knappheit durch Techniken wie Priorisierung, Scheduling oder Bandbreitenreservierung zu implementieren. Die Trends bei



den technischen Möglichkeiten sowie der Wirtschaftlichkeit von Netzwerkkomponenten begünstigen das Rightsizing.

Das Rightsizing unterscheidet sich von Overprovisioning dadurch, dass versucht wird, die Überlasten im Netzwerk auf extreme Fälle zu begrenzen. In diesen Netzen wird eine gewisse Zusatzkapazität über den üblichen Datenverkehr, sogenannte Datenspitzen, zur Verfügung gestellt. In der Praxis gleicht dieser Ansatz der bei der klassischen Telefonie verwendeten Erlang-Berechnung für die höchste Auslastung.

#### ▶ ANMERKUNGEN: RIGHTSIZING

- Die im Vergleich zur Überdimensionierung kostengünstigere, weil auf den Bedarf hin besser angepasste Realisierung hoher Bandbreiten im LAN erfordert detaillierteres Wissen über den Bandbreitenbedarf einzelner Anwendungen.
- Fehlertoleranzmechanismen und Monitoring müssen im Vergleich zur Überdimensionierung die Netzperformance engmaschiger erfassen, um rechtzeitig auf Überlasten reagieren zu können.
- Der »richtigen Dimensionierung« (Rightsizing) als Instrument der Kostenoptimierung im Bandbreitenmanagement sind im WAN Grenzen gesetzt. Bei Standardanschlüssen ist wegen fehlender Priorisierung eine gewisse Überdimensionierung unverzichtbar, für die Einrichtung von Priorisierungsmechanismen müssen zusätzliche Service-Level-Agreements abgeschlossen werden.

Beim Rightsizing geht es nicht ausschließlich um die Bandbreite. Die Vermeidung von Überlasten verringert die Paketverluste und den Jitter. Grundlage des Rightsizings ist eine optimierte Netzarchitektur, welche die Anzahl der Hops auf ein Minimum reduziert und Engpässe vermeidet.

Das Rightsizing wird hauptsächlich in kleineren und mittleren Unternehmensnetzen auf der Basis von Gigabit-Ethernet genutzt. Der Lösungsansatz des Rightsizings gehört zu den einfachsten und preiswertesten Wegen, einen NSQ im Netzwerk zu realisieren. Einen ähnlichen Ansatz verfolgen viele Service-Provider in ihren kommerziellen Backbones: Da die mittlere Auslastung dieser Backbones in der Regel bei 10 bis 15 Prozent liegt, werden Überlastungen in Spitzenzeiten umgangen.

WAN-Verbindungen werden von den Anbietern auf Basis der höchsten Auslastung beziehungsweise der Übertragungskapazität vermarktet. Die Unternehmen kaufen in der Regel reine Zugangsbandbreiten ein. Da diese Bandbreiten von den Service-Providern, unabhängig von deren Auslastung, immer mit dem oberen Durchsatz-Limit abgerechnet werden, entstehen auf Grund des Kostendrucks beim Übergang zwischen LAN und WAN natürlich Engpässe. Werden reine Datenapplikationen über die WAN-Verbindungen genutzt, haben die Bandbreitenbeschränkung und die eventuell entstehenden Engpässe kaum Auswirkungen.

Werden jedoch Sprach- und Videoanwendungen im Netzwerk eingesetzt, konkurrieren diese Datenströme um die gleiche Bandbreite im WAN und führen schnell zu einer weiteren Überlastung in den Netzsegmenten.

## 4. Quality-of-Service-Techniken

Netzwerke auf der Basis von Ethernet und IP sind von Hause aus eigentlich nicht für Echtzeitanwendungen wie Telefonie oder Video geeignet. Quality-of-Service-Techniken sollen IP-Netze echtzeitfähig machen und die Funktion von wichtigen Applikationen auch bei Lastspitzen sicherstellen.

Im ursprünglichen Design der TCP/IP-Protokolle spielte das Thema Quality-of-Service, kurz QoS genannt, eine untergeordnete Rolle. TCP/IP war für den reinen Transport von Anwendungsdaten, also Dienste wie File-Transfer, E-Mail oder Terminalzugriff, gedacht. Eine Übermittlung von Sprache oder Videos über ein TCP/IP-Netzwerk haben die Entwickler im ursprünglichen Designkonzept nicht vorgesehen.

Der Begriff Quality-of-Service beschreibt in der TCP/IP-Welt die Güte eines Kommunikationsdienstes aus der Sicht der Anwender. Dabei wird verglichen, wie stark die Güte des Dienstes mit deren Anforderungen übereinstimmt. Formal ist QoS eine Menge von Qualitätsanforderungen an das gemeinsame Verhalten beziehungsweise Zusammenspiel von mehreren Systemen.

Der Begriff Verkehrsgüte bezeichnet den von der jeweiligen Bemessung der Betriebsmittel abhängigen und quantifizierbaren Teil der QoS. Die Verkehrsgüte eines Netzwerks hängt von der Bemessung der Anzahl von Leitungen und Steuereinrichtungen ab. Quantitative Beschreibungen der Verkehrsgüte benutzen Parameter der Verkehrstheorie wie Wartewahrscheinlichkeit, mittlere Wartedauer oder Verlustwahrscheinlichkeit.

QoS bezeichnet allgemein die Dienstgüte von Übertragungskanälen. Durch Quality-of-Service werden bestimmte Parameter aktiv beeinflusst, die für das Management der Network-Service-Quality verantwortlich sind. Hierzu gehören Bandbreite, Verzögerung, Paketverluste, Jitter und die Priorisierungen. Durch QoS lassen sich die wichtigen Parameter der Sprachqualität indirekt beeinflussen. QoS basiert auf definierten Kontrollmechanismen im Zusammenspiel mit einem Verkehrsmanagement, um die NSQ-Metriken des Transportnetzes in Bezug auf die übermittelten Informationsströme zu beeinflussen.

QoS ist somit kein zusätzliches Feature, das sich an eine Netzinfrastruktur wie ein Add-on anflanschen lässt. QoS ist vielmehr das Resultat einer Vielzahl von aufeinander abgestimmten Maßnahmen, die im grundsätzlichen Design einer Netzinfrastruktur verankert sein müssen. Um zu verstehen, wie QoS entsteht, ist es nützlich sich zu vergegenwärtigen, dass Netzinfrastrukturen einer Architektur in Schichten folgen.

Das Augenmerk zum Erreichen von QoS muss also auf das Zusammenwirken der verschiedenen Schichten gelenkt werden. Hieraus kann ermittelt werden, welche Anforderungen sich für die verschiedenen Schichten ergeben. Im Grundzug



kann davon ausgegangen werden, dass sich die Quality-of-Service im Schichtenmodell »nach oben vererbt«. Es ist jedoch immer nützlich zu überlegen, ob in einer Schicht Mechanismen möglich sind, die eine Toleranz gegen Mängel von unterliegenden Schichten erlaubt. Das Schichtenmodell macht deutlich, dass nur eine Menge von abgestimmten Maßnahmen in den verschiedenen Schichten zu einem Gesamtpaket führt, das QoS zu einer integrierten Eigenschaft werden lässt.

Die QoS-bezogene Priorisierung oder Parametrisierung des Datenverkehrs, die Datenratenreservierung, die Datenratenlimitierung und die Paketoptimierung wird durch folgende Mechanismen realisiert:

- ▶ VLANs (Schicht 2),
- ▶ IEEE 802.1p/Q (Schicht 2),
- ▶ Differentiated-Services, kurz DiffServ, (Schicht 3),
- ▶ Integrated-Services, kurz IntServ, (Schicht 3),
- ▶ Traffic-Shaping (Schicht 2 bis Schicht 7) sowie
- ▶ Quality-of-Experience, kurz QoE, (Schicht 3 bis Schicht 7).

#### ▶ ANMERKUNGEN: VLANS (IEEE 802.1Q)

■ VLANs ermöglichen eine kostengünstige Realisierung separater Kommunikationskanäle im LAN auf Ebene 2. Insbesondere in einfach strukturierten Netzen ist diese Methode in der Praxis bewährt. Detaillierteres Wissen über den Bandbreitenbedarf einzelner Anwendungen ist zwar erforderlich, anspruchsvolleres Know-how für Routing-Konzepte auf Ebene 3 ist dafür jedoch nicht erforderlich.

■ Ohne weitere Mechanismen wie IEEE 802.1p wird mit der Einrichtung von VLANs allein noch keine Sicherung von verfügbaren Bandbreiten für Sprachdienste erreicht, da sich das VLAN die verfügbare Bandbreite mit allen Kanälen im LAN teilt.

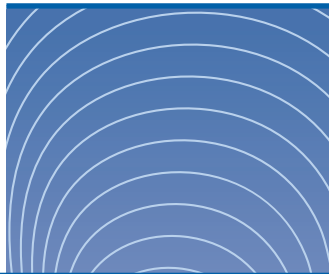
■ Die Einrichtung von VLANs nach IEEE 802.1Q hat keine Auswirkung auf die Bandbreitenverfügbarkeit an den Schnittstellen zum WAN.

## 4.1 Virtuelle LANs

Die Integration von virtuellen LANs, kurz VLANs, ist die einfachste Methode zur Verkehrssteuerung. Der Standard 802.1Q definiert die Architektur und die Dienste in Virtual-Bridge-LANs sowie die Protokolle, Algorithmen und Verfügbarkeit dieser Dienste. Dabei muss jedoch beachtet werden, dass es sich bei diesem Ansatz mehr um ein Sicherheitskonzept handelt, als um eine Verbesserung der Dienstqualität. Aus diesem Grund sind VLANs von IEEE 802.Q unabhängig. Der Grundstandard für Priorisierung ist IEEE 802.1p, der IEEE 802.1Q setzt auf diesem auf.

Das IEEE 802.1Q-Konzept wird in der Regel mit anderen Bandbreitenmanagementmechanismen wie den Differentiated-Services kombiniert. Durch VLANs werden die Sprachkomponenten exklusiv einem VLAN zugeordnet. Durch die VLAN-Technik wird ein Ethernet-Netzwerk in mehrere virtuelle Netzwerke aufgeteilt. Über ein bestimmtes VLAN können dann beispielsweise nur die Telefondatenströme Ende zu Ende übermittelt werden. Die VLANs enden in der Regel an den WAN-Grenzen des lokalen Netzwerks. Daher muss am Übergang ins WAN mit entsprechenden Mechanismen das VLAN in die betreffenden Mechanismen des WLANs umgesetzt werden. Beispielsweise werden in der Praxis die VLANs auf die MPLS-Kanäle/VLANs gemappt. Außerdem müssen an den Netzgrenzen mit Hilfe von SLAs die individuellen Parameter gesichert werden.

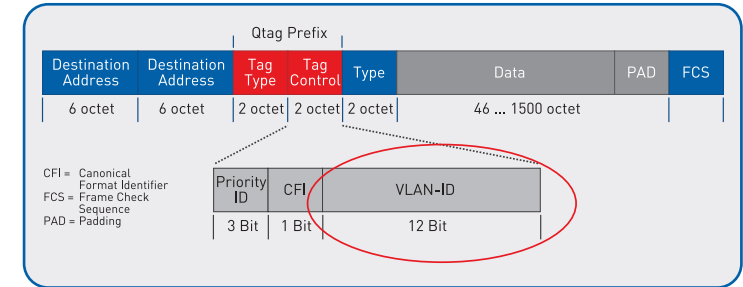
Mit einem VLAN steht ein paralleles logisches Netzwerk zur Verfügung. In diesem Netzwerk konkurrieren die aktiven Geräte um die verfügbare Bandbreite. Aus diesem Grund müs-



sen auch in einem VLAN die verfügbaren Bandbreiten genau kontrolliert und überwacht werden.

### 4.2 IEEE 802.1p/Q

Bridges (Brücken) verbinden in einem lokalen Netzwerk zwei oder mehrere Netzsegmente auf dem MAC Layer (Schicht 2a). Der Standard IEEE 802.1D beschreibt die grundsätzlichen Funk-



Tagged-Frame nach IEEE 802.1p/Q

Grafik: VAF

#### PARAMETER

#### BESCHREIBUNG

Service-Availability	Die Dienst-Verfügbarkeit entscheidet über die Bereitstellung eines Dienstes in Bridged-LANs.
Frame-Loss	Das Äquivalent zu Paketverlusten.
Frame-Misorder	Innerhalb des MAC-Levels sind Neuansorderungen von Paketen/Frames nicht möglich, deshalb ist die Reihenfolge der Pakete ein QoS-Parameter.
Frame-Duplication	Das Duplizieren von Paketen wird von MAC-Services nicht unterstützt.
Transit-Delay experienced by Frames	Die Zeit, die zwischen der Anforderung und der Bestätigung eines erfolgreichen Transfers verstreicht.
Frame-Lifetime	Die Lebensdauer eines Paketes innerhalb eines Netzwerkes wird begrenzt, um mögliche Schleifenbildungen vorzubeugen. Dies entspricht dem Time-to-Live-Wert im Header, kurz TTL.
Undetected-Frame-Error-Rate	Die Fehlerrate ist sehr gering, wenn die Verfahren zur Ermittlung von Check-Summen, Frame-Checksum oder kurz FCS, zur Anwendung kommen.
Maximum-Service-Data-Unit-Size supported	Die maximale Größe eines Paketes innerhalb eines Netzwerkes wird von demjenigen Netzelement vorgegeben, welches die kleinste Paketgröße vorgibt.
User-Priority	Innerhalb des MAC-Level wird die User-Priorität als QoS-Parameter behandelt.
Throughput	Der Datendurchsatz.

Die QoS-Parameter nach IEEE 802.1p

tionen von MAC-Bridges. Diese Bridges teilen ein Netz in verschiedene Kollisionsdomänen auf.

Da durch den Einsatz von MAC-Bridges jeder angeschlossene Netzstrang nur die Pakete empfängt, deren Empfänger sich auch in diesem Netz befindet, vermindert sich die Netzlast in den jeweiligen Netzsegmenten. MAC Layer Bridges verfügen über einen Selbstlernmechanismus. Mit dessen Hilfe lernt die Bridge die Absender-MAC-Adressen von Paketen der einzelnen Teilnetze. Diese werden automatisch in eine interne Weiterleitungstabelle eingetragen. Aus diesem Grund werden diese Bridges auch als transparente Bridges bezeichnet. Anhand der gelernten MAC-Adressen bestimmt die Bridge den Weg zum Empfänger. Die Absenderadressen werden laufend aktualisiert, um Änderungen sofort zu erkennen.

In der Fachliteratur gibt es keine eindeutige Unterscheidung zwischen Bridges und Switches. Bei Layer 2 Switches handelt es sich um eine Weiterentwicklung der Bridging-Technik. Switches arbeiten als transparente Bridges, haben jedoch eine höhere Durchsatzleistung und mehr Ports.



Zur Sicherung der Dienstgüte innerhalb des MAC-Layers (Layer 2a) wurde der Standard der MAC Bridges (IEEE 802.1D) um die Substandards IEEE 802.1p und IEEE 802.1Q ergänzt. Der IEEE 802.1Q Standard definiert eine Architektur für »Virtual Bridged LANs« und dessen Services. Dieser herstellerübergreifende Substandard beschreibt einen Mechanismus zur Nutzung eines gemeinsamen physikalischen Netzsegments durch mehrere virtuelle Netzwerke. Hierzu werden zusätzliche Datenfelder für das VLAN-Tagging festgelegt. Diese VLAN-Tags werden in das betreffende Ethernet-Paket eingefügt. Ein VLAN-Tag besteht aus vier Feldern mit einer Gesamtlänge von 32 Bit.

Der IEEE 802.1Q Standard enthält keinerlei Festlegungen zum Bandbreitenmanagement auf der Schicht 2. Diese Lücke schließt der IEEE 802.1p (»Traffic Class Expediting and Dynamic Multicast Filtering«) Substandard. Dieser beschreibt Methoden für die Bereitstellung der Dienstgüte auf dem MAC-Level.

		Anzahl verfügbarer Verkehrsklassen							
		1	2	3	4	5	6	7	8
User-Priorität	0	0	0	0	1	1	1	1	2
	1	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	1
	3	0	0	0	1	1	2	2	3
	4	0	1	1	2	2	3	3	4
	5	0	1	1	2	3	4	4	5
	6	0	1	2	3	4	5	5	6
	7	0	1	2	3	4	5	6	7

Mappen der User-Priority in die Verkehrsklassen

Der Standard IEEE 802.1p definiert die Werte innerhalb des Priority-Felds. Die Werte Tag-Protocol-Identifier (TPID): 0x8100 und Canonical Format Identifier (CFI): 0x0 (Ethernet), sind fest definiert. Das VID-Feld (VLAN Identifier) wird vom Standard IEEE 802.1Q benutzt und dient zur Identifizierung der virtuellen LANs (VLANs).

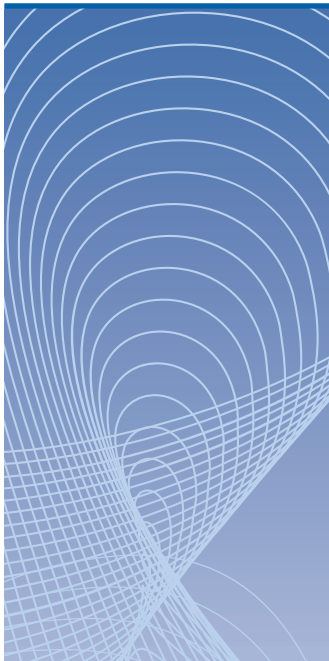
Priorität	Bitmuster	Typ	Dienstklasse
0	000	BE	Best Effort, keine Priorisierung
1	001	BK	Background, Hintergrunddienste (Newsticker)
2	010	-	Reserviert
3	011	EE	Excellent Effort, Allgemeine Datendienste
4	100	CL	Controlled Load, Steuerdienste (Routing)
5	101	VI	Video
6	110	VO	Voice, Sprachdaten
7	111	NC	Network Control, Höchste Priorität für Netzkontrolle

Unterscheidung der unterschiedlichen Verkehrsarten

Die Priorisierung der Pakete und die entsprechenden Zuordnungen zu den Queues im Switch werden durch das Prioritätsfeld im VLAN-Tag (Bits 1 bis 3) erreicht. Der IEEE 802.1p-Standard legt das Mappen der Prioritäten zu den jeweiligen Verkehrsklassen (Traffic-Classes) und damit zu den vorhandenen Queues fest. In der nachfolgenden Tabelle wird dies dargestellt.

Innerhalb eines Netzwerks/VLANs werden die transportierten Informationen in verschiedene Verkehrsarten oder Traffic-





User-Priority-Wert	Verkehrsklasse, Bemerkung	
0	Best-Effort	normaler Datenverkehr
1	Background	großvolumige Datensendungen, die zeitunkritisch sind und keine besondere Priorität erhalten
2	Spare	Standarddaten
3	Excellent-Effort	wichtige Daten
4	Controlled-Load	zeitkritische Daten mit hoher Priorität
5	Video	Bildübertragung mit einer Verzögerung < 100 ms
6	Voice	Sprachübertragung mit einer Verzögerung < 10 ms
7	Network-Control	Netzwerk-Management-Daten, Daten, die dazu dienen, den Netzbetrieb aufrecht zu erhalten, Daten, die der Ermittlung der kürzesten Route dienen, wie bei OSPF

Unterscheidung der Verkehrsklassen in verschiedene Verkehrsarten

**ANMERKUNGEN: IEEE 802.1p**

- Die Priorisierung erfolgt auf Ebene 2 und stellt keine Anforderungen an die höheren Protokollschichten. Der zusätzliche Protokoll-Overhead ist mit vier Byte pro Paket vernachlässigbar.
- Das Verfahren ergänzt Ansätze der Dimensionierung und der Strukturierung (VLAN) von Bandbreiten um statische Priorisierungsregeln für zugeordnete MAC-Adressen. Bei hoher Last zeigen die niedriger priorisierten VLANs/Geräte nach definierten Vorgaben ein verzögertes Antwortverhalten. Das Verfahren liefert dadurch einen Beitrag zur (kosten-)optimierten Bereitstellung »richtig dimensionierter« WAN-Strecken.
- Die Priorisierung erfordert detailliertes Wissen über den Bandbreitenbedarf einzelner Anwendungen, Fehlerortung und Fehlerbehebung.

Types unterschieden. In der Regel wird die jeweilige Priorität des Pakets wird durch die Anwendung oder durch die Berechtigung des Nutzers definiert.

Die Unterscheidung von sieben Verkehrsarten im Gegensatz zu acht definierten Verkehrsklassen verringert die Anzahl der benötigten Queues im Layer 2 Switch.

Tipp für die Praxis: Dedizierte Sprach-VLANs werden auf Dauer eher abgeschafft bzw. werden in den Sprach-VLANs zukünftig vermehrt auch Daten übertragen und in Daten-VLANs auch Sprache übertragen. Durch die steigende Verbreitung von Softphones und Smartphones geht die eindeutige Zuordnung von Endgeräten und Applikationen zu speziellen VLANs verloren, so dass dadurch auch der Verkehr in den VLANs konvergiert.

### 4.3 Integrated-Services

Die Grundidee der Integrated-Services, kurz IntServ, ist ähnlich der Bandbreitenfunktionen in Telefonnetzen. Ziel der Integrated Services ist es, eine feste Bandbreite auf den Übertragungstrecken für einzelne Teilnehmer zu reservieren und eine entsprechende Verzögerung einzuhalten. Somit ist IntServ ein feingranulares QoS-Verfahren zur Priorisierung von IP-Datenströmen. Im Gegensatz zu anderen QoS-Verfahren werden die Ressourcen für die jeweiligen Verbindungen und nicht für Verkehrsklassen angefordert.

Damit stellt IntServ theoretisch einen sehr effektiven Mechanismus zur Verfügung, der sich jedoch in der Netzwerkpraxis nur sehr eingeschränkt etablieren konnte. Gründe hierfür dürften in der Komplexität, schwierigen Administrierbarkeit und



den erforderlichen, enormen Rechnerleistungen zu sehen sein. Da wegen der theoretischen Leistungsfähigkeit IntServ in Diskussionen immer wieder angeführt wird, werden die Grundzüge nachfolgend in gebotener Kürze dargestellt.

Bei IntServ fragt der Sender bei den Routern im lokalen Netz oder im Internet an, ob die Ressourcen für einen Verkehrsfluss vom Sender zum Empfänger vorhanden sind. Ist das der Fall, stellen die Vermittlungssysteme die Verbindung in der gewünschten Qualität her und halten sie so lange aufrecht, bis die Übertragung zu Ende ist.

IntServ definiert Serviceklassen zur Realisierung einer definierten Dienstgüte. IntServ arbeitet mit Reservierungen, die über das Ressource-Reservation-Protocol (RSVP) festgelegt werden. IntServ und RSVP sind jedoch getrennte Konzepte, da IntServ auch andere Reservierungsmechanismen verwenden könnte und andererseits RSVP mehr Signalisierungsmöglichkeiten bietet als für IntServ benötigt werden. Bisher wurden nur zwei der existierenden Serviceklassen in Verbindung mit RSVP festgelegt:

- ▶ **Der Guaranteed Service** verlangt, dass eine maximale Ende-zu-Ende-Übertragungszeit eingehalten wird. Daher muss die Übertragungszeit über eine virtuelle Verbindung nach oben beschränkt sein. Diese Serviceklasse wird von Anwendungen wie IP-Telefonie und Videoübertragungen für die Übermittlung der Daten über Netzwerke benötigt.
- ▶ **Der Controlled Load Service** ermöglicht die Zuweisung bestimmter Bandbreiten für einzelne Anwendungen. Alle ande-

ren QoS-Parameter bleiben dabei unberücksichtigt. Eine Dienstgüte gemäß den Definitionen der Controlled Load erfordert keine zusätzlichen Mechanismen auf den unteren Ebenen des ISO/OSI-Referenzmodells und wird von Anwendungen in Anspruch genommen, die tolerant gegenüber geringen Paketverlusten und Delay sind.

Integrated-Services fordert vom Sender eine Datenflussspezifikation (flowspec). Diese dient der Abarbeitung der vereinbarten Charakteristiken zur Weiterleitung und Übermittlung der Pakete:

- ▶ **Traffic-Spezifikation (Tspec)** beschreibt die Variation der angeforderten Bandbreite über der Zeit. Zur Beschreibung der Tspec wird das Konzept des Token-Bucket-Filters verwendet. Hierzu stelle man sich einen Behälter mit einem bestimmten Volumen vor. Von einem Tokengenerator werden Token mit der konstanten Rate  $r$  erzeugt. Damit wird der Behälter (Bucket) kontinuierlich gefüllt. Kommt ein Datenpaket an, so muss die Anzahl der Token im Bucket mindestens so groß wie die Byte-Anzahl des Pakets sein. Ist das der Fall, wird das Paket als filterkonform weiter verarbeitet und der Inhalt des Buckets um die Größe des Datenpakets verringert. Sonst wird das Paket als nicht filterkonform behandelt. Dabei ist die Behandlung dieser Daten je nach Anwendungsfall sehr unterschiedlich. Die Daten können verworfen oder im Best-Effort-Modus übertragen werden. Kommen über einen längeren Zeitabschnitt keine Datenpakete an, dann füllt sich der Bucket. Ist der Bucket voll, werden neu erzeugte Token verworfen.

#### Parameter Tspec

p	Spitzen-Datenrate in Byte/s (peak rate of flow)
b	Bucket-Tiefe in Byte
r	Token-Bucket-Rate in Byte/s
m	minimale unterstützte Paketgröße in Byte
M	maximale Paketgröße in Byte

► **Request-Spezifikation** (Rspec) beschreibt die vom Netzwerk verlangte Dienstgüte. Die Angabe lautet Controlled-Load – ohne weitere Parameter – oder Guaranteed-Service mit Angabe einer maximalen Verzögerung.

#### Parameter Rspec

R	Bandbreite in Byte/s
S	Slack-Terms

Durch die Reservierung der notwendigen Bandbreite und die gleichzeitige Festlegung der Spitzen-Datenrate (p) werden geringe Paketverluste garantiert. Die Parameter (r, b) zur Bestimmung des Token-Bucket-Verhaltens sorgen zudem für eine niedrige Laufzeitverzögerung. Der Parameter S (Slack-Term) wird generell auf den Wert = Null gesetzt.

Guaranteed-Service und Controlled-Load-Service können nur in Netzwerken mit geregelten Bandbreiten genutzt werden. Dabei stehen die in der Verkehrscharakteristik (Tspec; Traffic-Specification) und in der Reservierungscharakteristik (Rspec; Reservation-Specification) vereinbarten Parameter im direkten Zusammenhang mit den Service-Level-Agreements (SLAs). Diese Parameter werden mit Hilfe der Admission/Policy-Kontrolle an die Layer 3 Komponenten (Router) übermit-

telt. Aus diesem Grund benötigt IntServ leistungsfähige Netzzelemente die folgenden Funktionen unterstützen:

► **Admission-Control:** Jedes IntServ-Endgerät sowie jeder IntServ-fähige Router muss über eine Admission-Control-Instanz verfügen. Empfängt ein IntServ-Gerät eine Reservierungsanfrage, prüft die Admission-Control, ob im Gerät genügend Ressourcen zur Verfügung stehen und ob der gewünschte Dienst mit dem auf dem zugehörigen Interface eingesetzten Paket-Scheduling-Verfahren gewährleistet werden kann.

► **Classification:** Sorgt für die Erkennung von Paketen, die bestimmte QoS-Anforderungen stellen, beziehungsweise zu bestehenden Reservierungen gehören.

► **Policy-Control:** Die Policy-Control-Funktionseinheit sorgt für die Authentisierung der Teilnehmer im Netz und prüft die Berechtigungen der Teilnehmer in Bezug auf deren QoS-Anforderungen.

► **Queuing and Scheduling:** Organisiert die Warteschlangen damit die zugesagten Dienstgüten erreicht werden. Jede Warteschlange erhält auf Grund ihrer Priorität einen festen Anteil der verfügbaren Bandbreite.

► **Das Resource-Reservation-Protocol**, kurz RSVP, ermöglicht die Reservierung von Ressourcen zur Sicherstellung der Dienstqualität bei der Datenübertragung. Die dazu benötigten Informationen speichern und verwalten die Router in sogenannten State-Blocks.

Der Schlüssel für Echtzeitübertragungen im Internet liegt in der Reservierung von Bandbreite entlang des Ende-zu-Ende-Kommunikationspfades. RSVP hat die Aufgabe, diese Reservierungen zu realisieren und damit Probleme durch Verzögerungen, Paketverluste und fehlende Bandbreite zu verhindern.



Auf Basis des RSVP-Protokolls werden die individuellen Reservierungsanforderungen an die Netzknoten weitergeleitet. Für die Sicherstellung der Dienstqualität sind jedoch die Netzknoten verantwortlich.

RSVP arbeitet empfängerorientiert. Aus diesem Grund sind immer die Empfänger entlang des jeweiligen Kommunikationspfads für die Reservierungen verantwortlich. Reservierungsanforderungen können abgelehnt werden. Um Bandbreite einzusparen, können kompatible Anforderungen zu neuen Anforderungen vereinigt werden. RSVP reserviert Ressourcen nur in einer Richtung. Die unidirektionale Reservierung erfordert, dass bidirektionale Anwendungen (beispielsweise VoIP) zwei voneinander unabhängige Reservierungen etabliert werden. Die Reservierungen müssen in regelmäßigen Abständen bestätigt werden, sonst werden die reservierten Ressourcen wieder freigegeben.

Das Grundschema von RSVP sieht keine Rückmeldungen an den Empfänger vor. Der Empfänger schickt eine Anforderung in Richtung Sender. Die Netzknoten akzeptieren diese oder leh-

nen sie ab. Im Fall einer Ablehnung erhält der Empfänger eine Fehlermeldung. Wird die Anforderung akzeptiert, hat der Empfänger keine genaue Kenntnis über die Qualität der Verbindung. Eine Erweiterung zum RSVP-Grundschema wird als »One Pass with Advertising« bezeichnet. Entlang des jeweiligen Kommunikationspfads werden dazu Kontrollpakete in Richtung des Empfängers geschickt und dadurch die Qualität der aktuellen Verbindung ermittelt.

Bei einer Routenänderung erzeugen die nächsten Path- und Resv-Messages entlang der geänderten Route neue Kontrollzustände in den Netzknoten.

In klassischen Telefonnetzen reserviert das leitungsvermittelte Netzwerk zu Beginn einer Verbindung die notwendige Bandbreite. Diese Reservierungen werden bis zum Ende der Verbindung bereitgestellt. Im Gegensatz dazu, nutzt der RSVP-Mechanismus zur erforderlichen Ressourcenreservierung dynamische Kontrollzustände. Die Kontrollzustände werden automatisch entfernt, wenn diese nicht regelmäßig Refreshes – in Form von Path- und Resv-Messages – bestätigt werden. Durch explizite Teardown-Messages lassen sich die Reservierungen entfernen.

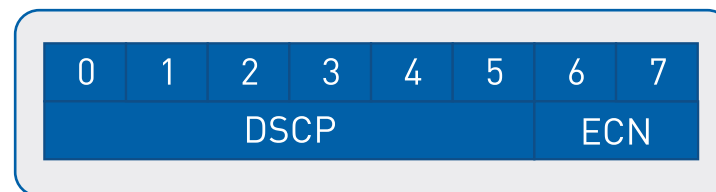
Auch in Netzwerkbereichen ohne RSVP-Unterstützung muss die Übertragung der Daten sichergestellt sein. Auf diesen Teilstrecken ist in der Regel keine Reservierung von Ressourcen möglich, so dass die Einhaltung der vom Empfänger geforderten Dienstqualität davon abhängt, ob der zu überbrückende Bereich zufällig über ausreichend Ressourcen verfügt, um die Dienstqualität auch ohne Reservierung sicherstellen zu können. Generell steht auf solchen RSVP-losen Teilstrecken nur die Best-Effort-Übermittlung zur Verfügung.

#### ▶ ANMERKUNGEN: INTSERV

- Die Applikationen müssen den Intserv-Mechanismus unterstützen. Dies ist in den meisten aktuellen Produkten für private Netze nicht der Fall, weshalb dieses Verfahren heute kaum noch eingesetzt wird.
- Das Verfahren beschränkt sich weitgehend auf den Einsatz der ATM-Technologie im Backbone-Bereich öffentlicher Telekommunikationsnetze. Unter anderem wegen der schwierigen Administrierbarkeit konnte ATM und damit Intserv sich nicht im Bereich Privater Netze durchsetzen.

## 4.4 Differentiated-Services

Beim IP-Protokoll (RFC 791) auf Schicht 3 erfolgt die Datenübermittlung ungesichert und verbindungslos. Alle Pakete werden vom IP unabhängig von den anderen übertragen. Zusätzliche Funktionen wie Fehlererkennung, Überwachung der Reihenfolge, Flusskontrolle und Sicherung der Übertragung werden durch die Protokolle höherer Schichten realisiert. IP-Netze unterscheiden somit nicht zwischen verschiedenen Anwendungen und der Priorität der über das Netzwerk übermittelten Datenpakete. Diesem Problem kann man durch Bereitstellung hoher und teurer Übertragungskapazitäten begegnet werden. Das ursprüngliche IP-Protokoll definiert ein Type-of-Service-Feld (ToS) im Header, welches die Vorrangsteuerung, Wartezeit, Durchsatz und Zuverlässigkeit der jeweiligen Datagramme beschreibt. Somit wird im Gegensatz zu an-



DiffServ-Feld nach IETF

deren QoS-Verfahren die Priorität eines IP-Paketes bereits vom Sender bestimmt. Die Router auf dem Weg zum Empfänger entscheiden allein anhand der ToS-Angabe über die bevorzugte Weiterleitung zum Empfänger. Die ToS-Mechanismen wurden jedoch nie zufriedenstellend in die Praxis umgesetzt. Zur Realisierung einer skalierbaren QoS-Lösung mit unterschiedlichen Dienstklassen hat die IETF vor 10 Jahren den ToS-Mechanis-

mus durch die Differentiated-Services, kurz DiffServ, ersetzt. Die DiffServ-Funktionen sind in den RFCs 2474 und 2475 festgelegt.

Die Priorität eines IP-Paketes wird bereits vom Sender im IP-Header bestimmt. Die Koppelkomponenten auf dem Weg zum Empfänger entscheiden allein anhand dieser Angabe über die bevorzugte Weiterleitung zum Empfänger. Im Unterschied zum ToS-Verfahren kann durch die DiffServ-Mechanismen jedes IP-Paket von den Übermittlungskomponenten auf dessen Wichtigkeit geprüft werden. Bei Auftreten einer Überlast können unwichtigere Pakete verworfen werden, um eine priorisierte Zustellung der wichtigeren Pakete zu ermöglichen.

DiffServ verwendet dazu das ehemalige IPv4-TOS-Feld beziehungsweise das IPv6-Traffic-Class-Feld des IP-Headers, welches in ein DiffServ-Feld (DS) umdefiniert wird. Dieses besteht aus einem sechs Bit langem Differentiated-Service-Codepoint-Feld (DSCP) für 64 mögliche QoS-Klassen, wobei die restlichen zwei Bits von aktuellen TCP-Stacks zur Explicit-Congestion-Notification, kurz ECN, genutzt werden.

DiffServ benötigt keine Signalisierung und Zustandsspeicherung in den Koppelkomponenten, da jedes einzelne Paket individuell nach seiner Dienstklasse von den Netzelementen behandelt wird. Stattdessen definiert DiffServ eine kleine Anzahl von Regeln für Qualitätsklassen, so genannte Per-Hop-Behaviours oder PHBs, die die Router durch geeignete Maßnahmen unterstützen müssen.

DiffServ definiert die folgenden Qualitätsklassen.

► **Expedite-Forwarding (EF, RFC 3246):** Diese Qualitätsklassen mit strikter Priorität über alle anderen Klassen bieten den geringsten Paketverlust, sehr geringe Verzögerung und kleine Jit-

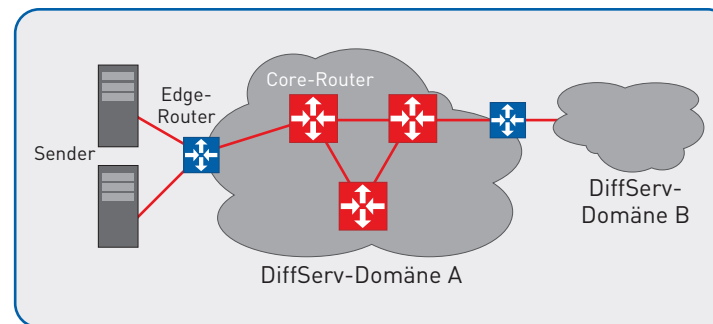
ter. Jeder Router reserviert einen bestimmten Prozentsatz seiner Kapazität für diesen Verkehr. Für die Anforderungen von VoIP-Verkehr in Richtung geringer Verzögerung eignen sich diese Qualitätsklassen ganz besonders.

► **Assured-Forwarding (AF, RFC 2597):** Innerhalb dieses Dienstes werden vier unterschiedliche Qualitätsklassen und drei Levels für den Paketverwurf pro Qualitätsklassen definiert. AF belegt daher zwölf DSCP-Werte und dient für Verkehr, der sensitiv auf Paketverluste reagiert.

► **Best-Effort (BE):** Es gibt keine Dienstgarantie, zur Verfügung steht die verbleibende Bandbreite.

Für die Sicherung der Dienstgüte – und somit für die unterschiedlich priorisierte Weiterleitung der Pakete – sorgen in einem DiffServ-fähigen Netzwerk die DiffServ-fähigen Koppelkomponenten. Die jeweiligen DiffServ-Definitionen gelten jedoch nur für ein zusammenhängendes administratives Netzwerkkonstrukt. Aus diesem Grund werden DiffServ-fähige Netze als DiffServ-Domänen bezeichnet.

Die folgende Abbildung zeigt die DiffServ-Architektur.



DiffServ-Architektur nach IETF

Grafik: VAF

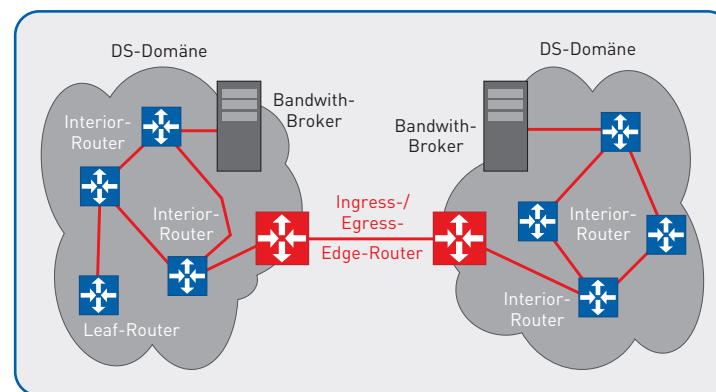
Wie in der Abbildung links ersichtlich setzt der Sender über die Service-Primitives der Applikation das DS-Feld im IP-Header.

Da der Sender die Priorität eines Paketes festlegen kann, entstehen Sicherheitsrisiken und unter Umständen Probleme bei der Einhaltung der Dienstgüte des Netzes. Aus diesem Grund überprüft der Edge Port im Zugangs-Switch die Einhaltung der Policies und korrigiert diese gegebenenfalls. Dadurch bleibt die Kontrolle der Dienstgüte beim Betreiber des Netzes und nicht beim Endgerät bzw. Eindringling.

Im Edge-Router erfolgt somit die Klassifizierung, indem bei jedem Paket das DS-Feld ausgewertet beziehungsweise durch ein vom Administrator vorgegebenes DSCP-Feld ersetzt wird. Dieser »Stempel« legt die weitere Behandlung des IP-Paketes innerhalb der betreffenden DiffServ-Domäne fest. Die Werte des DSCP-Felds weisen auf die dazugehörigen PHBs. Die Core-Router sind nur für das Routing/Weiterleitung der DiffServ-Pakete anhand der sechs DSCP-Bits zuständig, das heißt sie schließen anhand des DSCP-Stempels auf die weitere Behandlung des Pakets. Bei DiffServ beschränkt sich daher die Intelligenz auf den Edge, während Core-Router lediglich das DiffServ-Feld auswerten müssen. Der DiffServ-Ansatz ermöglicht eine schnelle und zuverlässige Weiterleitung von priorisiertem Verkehr, aber nur so lange der priorisierte Verkehr im Rahmen der zugeteilten Ressourcen – beispielsweise der Bandbreite – bleibt. Außerdem darf der hoch priorisierte Verkehr nur einen Bruchteil des gesamten IP-Verkehrs ausmachen. Daher ist eine Zugangskontrolle, auch Admission-Control genannt, unumgänglich, die den priorisierten Verkehr limitiert.

Die Vorteile von DiffServ liegen in der guten Skalierbarkeit für große Netze und in der Interoperabilität von DiffServ mit IntServ und MPLS. Nachteilig an DiffServ ist, dass alle Router die DiffServ-Mechanismen »verstehen« müssen. Darüber hinaus erlaubt der DiffServ-Ansatz lediglich eine relative Bevorzugung gegenüber anderen Dienstklassen und sobald der Verkehr einer Dienstklasse zu hoch ist, verhält sich DiffServ wie der normale Best-Effort-Transport. Daher ist es wichtig, zusätzlich für genügend Bandbreite für den hoch priorisierten Verkehr zu sorgen. Obwohl die Dienstklassen (PHBs) definiert sind, gibt es keine Festlegungen für die einzelnen Werte und jede Domain kann die PHBs unterschiedlich behandeln.

Die Prozess-Organisation ist besonders im WAN-Umfeld im Sinne einer Ende-zu-Ende-Bereitstellung der DiffServ-Definitionen besonders wichtig. Da mehrere Institutionen an diesem Prozess beteiligt sind, ist eine formalisierte Ausgestaltung der Handlungsabläufe unabdingbar. Ohne eine Prozess-Organisation muss ständig mit der Gefahr gerechnet werden, dass durch unangemessene oder falsche Maßnahmen die QoS einer Net-



Kommunikation zwischen DiffServ-Domänen

Grafik: VAF

zinfrastruktur empfindlich leidet. Innerhalb einer DiffServ-Domäne haben die darin arbeitenden Koppelkomponenten unterschiedliche Aufgaben.

Man unterscheidet folgende Gerätefunktionen:

▶ **Leaf-Router** verbinden das Netz eines Internet Service Providers (ISP) mit dem Netzwerk eines Unternehmens bzw. einer Organisation.

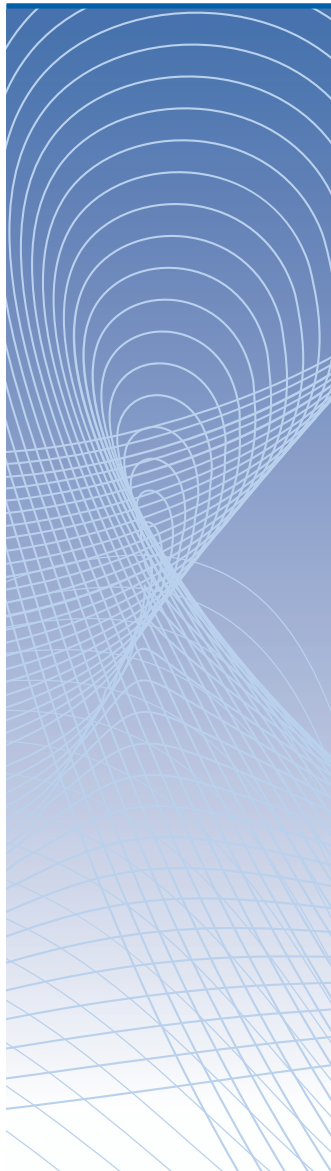
In der Regel weisen unterschiedliche Netzeinheiten verschiedene Priorisierungsrichtlinien auf. Die Anpassung an die Erfordernisse des Kunden erfolgt in der Praxis mit Hilfe von Service-Level Agreements (SLAs). Mit einem SLA garantiert ein Service Provider einem Kunden einen bestimmten Umfang an Netzwerkservices, die nach einer Preisstruktur abgerechnet wird. Die SLAs enthalten Information über die Leistungen des Service-Providers, wie garantierte Bandbreiten und Berechtigungen der Nutzer oder maximale Datenraten, die Peak-Bit-Rate. Die SLAs werden auch als Verkehrsprofile bezeichnet.

▶ **Boundary-Nodes:** Bezeichnet Router, die die Grenzpunkte zwischen unterschiedlichen Netzwerken bzw. unterschiedlichen DiffServ-Domänen darstellen. Diese werden in Egress-Edge-Router (in Richtung des Ausgangs) und Ingress-Edge-Router (in Richtung des Eingangs) unterschieden.

▶ **Egress-Edge-Router** haben die Aufgabe, den ausgehenden Verkehr an das individuelle Verkehrsprofil anzupassen. Dieser Vorgang wird auch als Traffic Shaping bezeichnet.

▶ **Ingress-Edge-Router** markieren und klassifizieren die eingehenden Pakete anhand des vorgegebenen Verkehrsprofils der DiffServ-Domäne.

▶ **Interior Router** bzw. **Core-Router** haben die Aufgabe die Pakete nach den vorgegebenen PHBs weiterzuleiten.



▶ Innerhalb des Internets erfolgt der Transport auch über die Netzgrenzen der Provider hinaus. Um die Service-Level-Agreements über diese Netzgrenzen hinaus zu erhalten, werden Traffic-Conditioning-Agreements, kurz TCAs, abgeschlossen. Dieses sind Abkommen zwischen den Service-Providern.

Sollen innerhalb eines Unternehmens/einer Organisation mehrere DiffServ-Domänen miteinander verbunden werden, müssen auch zwischen den jeweiligen Domänen entsprechende QoS-Funktionen in den angeschlossenen Netzwerken bereitgestellt werden. Nur so kann garantiert werden, dass der Datenverkehr einer bestimmten Prioritätsgruppe oder einer bestimmten Anwendung eine höhere oder niedrigere Servicequalität erhält. Hier werden zwischen Egress- und Ingress-Edge-Router die Traffic-Conditioning-Agreements wirksam. Diese Abkommen handeln die Betreiber der jeweiligen DiffServ-Domänen untereinander aus.

Die Grenzknoten zwischen den unterschiedlichen DiffServ-Domänen bzw. zwischen dem Kunden und dem ISP haben die Aufgabe, einkommende Datenflüsse angemessen zu formen. Insbesondere wird der Datenstrom in Übereinkunft mit einer auf administrativer Ebene vereinbarten Dienstgüte gebracht, dem so genannten Traffic-Conditioning-Agreement (TCA). Anhand von Messungen wird entschieden, ob der Datenstrom gegenwärtig dem vereinbarten Profil entspricht (IN-Profil) oder die Grenzen des Profils überschreitet (OUT-OF-Profil). Die Grundlage der TCAs ist der Standard RFC 2475, welcher eine Regel zur Bestimmung der Profile eines Paketes bereitstellt. Die Überwachung der TCAs wird innerhalb der Router einer DS-Domäne implementiert. Zu den Elementen des Traffic-Conditionings gehören:

▶ **Klassifizierung (Classifier):** Im DiffServ-Modell wählt der Classifier die Pakete aus dem Netzwerk-Verkehrswert aus. Ein Verkehrswert besteht aus einer Paketgruppe mit identischen Informationen in den folgenden IP-Header-Feldern: Quelladresse, Zieladresse, Quell-Port, Ziel-Port und Protokollnummer. Der Classifier ordnet den Verkehrswert in Klassen an, die auf den Eigenschaften basieren, die der Netzadministrator einer DiffServ-Domäne definiert hat.

▶ **Messen (Metering):** Im DiffServ-Modell überwacht eine Messeinheit (Meter) die Übertragungsrate des Verkehrswerts auf Klassenbasis. Der Meter wertet aus, inwieweit die tatsächliche Flussrate der konfigurierten Rate entspricht, um das geeignete Ergebnis zu ermitteln. Basierend auf dem Verkehrswert-Ergebnis wählt der Meter eine geeignete Aktion. Dies kann beispielsweise das Senden des Pakets an eine andere Aktion oder das Verwerfen des Pakets sein. Die Messmodule bestimmen, ob ein Datenfluss im Netzwerk der in der QoS-Richtlinie für diese Klasse definierte Übertragungsrate entspricht.

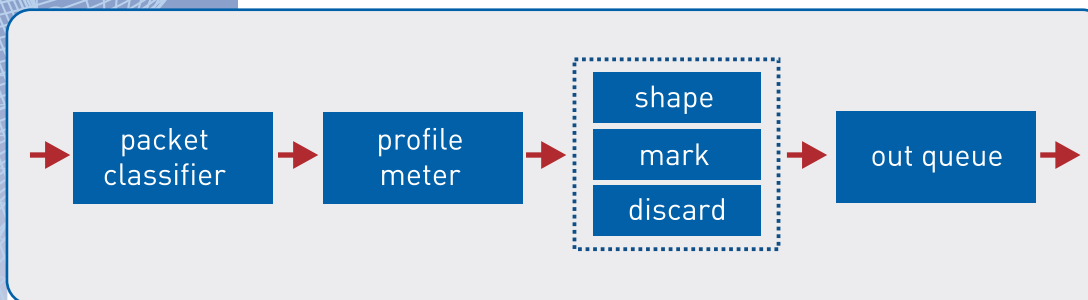
▶ **Markierung (Marking):** Das DiffServ-Konzept nutzt einen Marker zur Kennzeichnung der Pakete, die das betreffende Weiterleitungsverhalten widerspiegeln. Markierung ist der Prozess, einen Wert in den Paket-Header einzufügen, um festzulegen, wie das Paket zum Netzwerk weitergeleitet wird. Beispielsweise wird das DS-Feld mit einem numerischen Wert im IP-Paket-Header gekennzeichnet, der als Differentiated Services Codepoint oder DSCP bezeichnet wird. Eine DiffServ-konforme Koppelkomponente nutzt den DS-Codepoint, um das geeignete Weiterleitungsverhalten für das Paket einzuleiten.

▶ **Verkehrs Anpassung (Shaping/Dropping).** Die Traffic Shaping-Mechanismen in DiffServ-konformen Koppelkomponenten



werden zur gewünschten Beeinflussung der Datenpakete eines Datenstroms genutzt. Das Ziel besteht darin, die der QoS-Richtlinie entsprechen den Prioritätsstufen einzuhalten. Die Formung eines Datenstroms besteht für gewöhnlich im Verzögern eines Paketes. Die Verzögerung dauert so lange an, bis das Versenden des Paketes eine festgelegte Bandbreite in einem gewählten, kleinen Zeitrahmen nicht überschreitet. Übersteigen die realen Verkehrswerte einer Prioritätsstufe die Ausgangskapazität der Koppelkomponente, können die Pakete entweder (entsprechend der Drop-Wahrscheinlichkeit) verworfen oder auf eine niedrigere Priorität herabgestuft werden.

Die nachfolgende Abbildung zeigt die Funktionsweise des Traffic-Conditionings.



Traffic-Conditioning nach IETF

Die Unterscheidung von Paketen mit und ohne Profile verdeutlicht die Notwendigkeit, nur Datenströme zu erfassen, die den Parametern der Service-Level-Agreements oder den Traffic-Conditioning-Agreements entsprechen. Pakete, denen kein Verkehrsprofil zugeordnet ist, werden verworfen oder nach Best-Effort übertragen.

Die Administration der Verkehrsflüsse innerhalb einer DS-Domäne übernehmen Bandwidth-Broker. Diese haben folgende Aufgaben:

- ▶ Kontrolle der Ressourcen innerhalb einer DS-Domäne,
- ▶ Administration der Reservierungsanforderungen in Abhängigkeit des TCA und der momentanen Netzbelastung,
- ▶ Konfiguration der Engress/Ingress-Edge-Routers (Boundary-Nodes) sowie
- ▶ Policy.

#### ▶ ANMERKUNGEN: DIFFSERV

- Die Klassifizierung der Datenströme in Prioritätsklassen erfolgt im Edge-Bereich, also im Netzzugang-Switch des LAN/WAN. Die Prioritätsklassen müssen Ende-zu-Ende auf Ebene 3 konfiguriert werden.
- TCP/IP-fähige Anwendungen unterstützen Type-of-Service-Mechanismen (ToS). Die ToS-Informationen werden auf Ebene 3 durch Diffserv verarbeitet und ermöglichen in gerouteten Netzen geräteunabhängig eine differenziertere Priorisierung beispielsweise von Sprachdatenpaketen. Das Verfahren liefert dadurch einen Beitrag zur (kosten-)optimierten Bereitstellung »richtig dimensionierter« WAN-Strecken.
- Diffserv erfordert detailliertes Wissen über den Bandbreitenbedarf einzelner Anwendungen, das Design gerouteter Netze sowie Fehlerortung und Fehlerbehebung.

### 4.4.1 Exkurs zu DiffServ: Verwurfregeln bei Überlast

DiffServ bezeichnet das einem DSCP zugeordnete Weiterleitungsverhalten als Per-Hop-Behavior (PHB). Das PHB definiert die Prioritätsstufe der Weiterleitung, die ein markiertes Paket in Relation zu anderem Datenverkehr in einem DiffServ-konformen System empfängt. Diese Prioritätsstufe legt maßgeblich fest, ob ein markiertes Paket von der Layer-3-Komponente weitergeleitet oder verworfen wird.

Das Ziel eines PHB besteht darin, einen bestimmten Teil an Netzwerkressourcen für eine Verkehrsklasse im angrenzenden Netzwerk bereitzustellen. Dieses Ziel wird anhand der QoS-Richtlinie erreicht. Hierfür definiert der Systemadministrator die DSCP-konformen Prioritätsstufen für Verkehrsklassen. Die DiffServ-Architektur definiert zwei Arten von Weiterleitungsverhalten: Assured Forwarding und Expedited Forwarding.

#### Assured Forwarding

Mit Hilfe der Assured Forwarding (AF) PHB legt ein Netzbetreiber für den Empfang von IP-Paketen am Rand seiner DS-Domäne bestimmte Weiterleitungsregeln fest. Die AF-PHB definiert vier verschiedene Klassen für das Weiterleitungsverhalten. Diese werden mit Hilfe des Markers festgelegt.

Das Assured Forwarding Modell garantiert eine Rückwärtskompatibilität zu dem früher genutzten IP Precedence Modell. Obwohl die sechs Bit des DSCP Felds 64 (26) unterschiedliche Markierungen abbilden könnten, nutzt das Assured Forwarding Modell momentan nur 12 Werte.

Das AF-Modell teilt das DSCP Feld in drei Bereiche ein. Die drei Bit der höchsten Ordnung definieren die AF-Klasse. Die nächsten zwei Bit repräsentieren die Drop Preference innerhalb der jeweiligen Klasse. Das letzte Bit des DSCP-Felds wird immer auf den Wert = 0 gesetzt. Daraus ergibt sich folgende Wertigkeit:

**AAA DD 0**

**AAA** = Assured Forwarding class

**DD** = Drop Preference

**0** = Reserviert. (wird dieses Bit auf den Wert = 1 gesetzt, handelt es sich um einen vom Administrator festgelegten Wert)

Die vier Verkehrsklassen des Assured Forwarding Modells sind wie folgt gekennzeichnet:

**Class 1** = 001

**Class 2** = 010

**Class 3** = 011

**Class 4** = 100

Wird das alte ToS genutzt, werden zur Festlegung der Klasse eines Pakets die drei höchstwertigen Bits des ToS-Felds genutzt. Jede Klasse verfügt über drei Drop Preferences (low, medium und high). Die Drop Preferences werden durch die nächsten zwei Bit des DSCP Werts bzw. des ToS Bytes festgelegt.

**00** = No Drop Preference  
 (Class Selector oder Benutzer definiert)

**01** = Low Drop Preference

**10** = Medium Drop Preference

**11** = High Drop Preference

Bei Bandbreitenengpässen am Ausgang (bzw. in der Output-Queue) der Layer-3-Komponente wird durch den Congestion Avoidance (WRED) Mechanismus der Datenverkehr anhand der festgelegten Drop Präferenzen verworfen und eine Überlast abgewendet.

Sind alle sechs Bits des DSCP Bereich auf den Wert 000000 gesetzt, dann entspricht dieser Wert dem Best Effort. Die Daten werden dabei nicht priorisiert.

Beginnt das DSCP Feld mit den Werten 001, gehört das Paket zur Forwarding-Klasse 1. Die nächsten beiden Bits definieren die Drop-Präferenz des Pakets. Sind die nächsten zwei Bits auf den Wert Null (00) gesetzt, wird das Paket als CS1 (Klasse Wahlschalter 1). Rückruf aus früheren Blogs, dann handelt es sich um den Class Selector 1 (cs1).

Durch den Class Selector wird die Abwärtskompatibilität mit dem älteren IP-Precedence-Modell garantiert. Entsprechen die Drop-Präferenz-Bits dem Wert 00, dann wurde vom Administrator die Drop-Präferenz für die Klasse festgelegt. Die vier Werte für die Klasse 1 lauten:

Bezeichnung	Binär	Dezimal
CS1	001 00 0	8
AF11	001 01 0	10
AF12	001 10 0	12
AF13	001 11 0	14

**AF11** signalisiert, dass das Paket zur Assurance Forwarding (AF) Klasse 1 gehört und über eine niedrige Drop-Präferenz (1) verfügt.

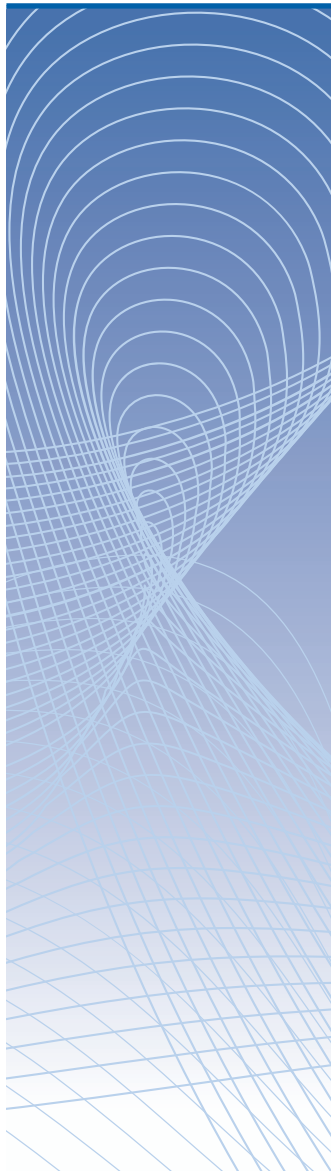
**AF12** gehört zur AF-Klasse 1 und verfügt über eine mittlere Drop-Präferenz (2),

**AF13** gehört zu AF-Klasse 1 und verfügt über eine hohe Drop-Präferenz (3).

Dies bedeutet für die Praxis: Bei Bandbreitenengpässen am Ausgang werden

- ▶ Pakete mit hoher Drop-Präferenz-Verkehr zuerst verworfen,
- ▶ danach wird der Verkehr mit einer mittleren Drop-Präferenz und
- ▶ zuletzt der Verkehr mit einer geringen Drop-Präferenz.

Da CS1 keine Drop-Präferenz enthält, werden diese Pakete nicht verworfen. Die Drop-Präferenzen der DSCP Bits stehen in engem Zusammenhang mit den Queuing-Mechanismen der Layer-3-Komponenten. Hauptsächlich wird heute hierfür der Weighted Random Early Detection (WRED) Mechanismus genutzt.



**Achtung:** Höhere Dezimalwert bedeuten nicht automatisch eine höhere Priorität. Der Schlüssel liegt in den Binärwerten zum Verständnis des Assured Forwarding Modells:

Bezeichnung	Binär	Dezimal
CS1	001 00 0	8
AF11	001 01 0	10
AF12	001 10 0	12
AF13	001 11 0	14
CS2	010 00 0	16
AF21	010 01 0	18
AF22	010 10 0	20
AF23	010 11 0	22
CS3	011 00 0	24
AF31	011 01 0	26
AF32	011 10 0	28
AF33	011 11 0	30
CS4	100 00 0	32
AF41	100 01 0	34
AF42	100 10 0	36
AF43	100 11 0	38

### Expedited Forwarding

Zur Kennzeichnung der Expedited Forwarding (EF) PHB ist der Wert = 101110 festgelegt. Dies entspricht dem DSCP 46 und

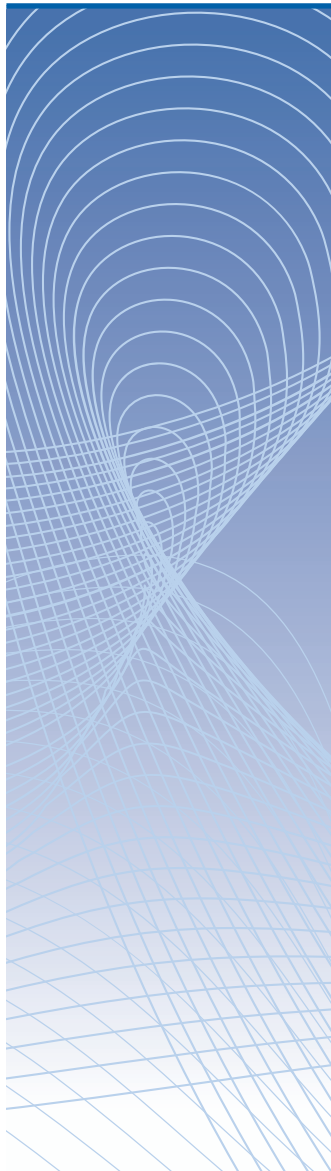
ist rückwärtskompatibel mit dem IP Precedence Wert = 5. EF folgt nicht den Regeln des Assured Forwarding Modells. EF stellt sicher, dass eine Verkehrsklasse mit EFs-bezogenem DSCP die höchste Priorität erhält. Verkehr mit einem EF DSCP wird nicht in eine Warteschlange gestellt. EF bietet geringen Verlust, Latenzzeit und Jitter. Das EF-Modell wird von den meisten VoIP-Herstellern als Default-Wert für die Übermittlung der Sprachdaten (RTP) genutzt. Der gesamte EF-Verkehr wird dabei von der Koppelkomponente mit Hilfe einer Priority Queue (PQ) übermittelt.

Die drei höchstwertigen Bits (101) werden nur berücksichtigt, wenn es sich um die Weiterleitung von IP-Precedence-Paketen handelt.

Das DSCP Feld nutzt nur die ersten sechs Bit (binär: 101110; dezimal: 46), während das ToS-Feld eine Länge von 8 Bit aufweist. Nimmt man die zwei fehlenden Bits (00) zum DSCP-Wert hinzu entsteht das binäre Gebilde 10111000.

Viele Netzmanagementanwendungen lassen sich vom Administrator nur dann konfigurieren, wenn dieser alle Bits des ToS Felds angibt. So benötigt der Befehl Ping-V auf einem Microsoft-Betriebssystem die Eingabe des gesamten ToS-Bytes. Auch der WireShark-Analysator stellt immer den gesamten ToS-Bereich dar, während Netflow das ToS-Byte im hexadezimalen Format abbildet.

Ein DSCP-Wert von 46 ergibt in der ToS-Darstellung einen Wert von 184. Obwohl man einem Ping den ToS-Wert von 184 zuordnen kann, wird der daraus resultierende ICMP- (Ping)-Verkehr wahrscheinlich nicht der hierfür vorgesehenen Anwendungsklasse zugeordnet werden.



Im RFC 4594 hat die IETF die QoS-Markierungen standardisiert. Die DSCP-Werte für die Applikationsklassen lauten:

- ▶ Best Effort
- ▶ Assured Forwarding Low Drop Prec (Class 4)
- ▶ Class Selector 5
- ▶ Expedited Forwarding (Per-Hop Behavior)

Für die Praxis empfiehlt sich im Zusammenhang mit RFC 4594 für DSCP und 802.1D die nachfolgenden Richtwerte einzustellen. Diese praxiserprobten Einstellwerte verlangen vom Administrator bei der Vielzahl in den Netzwerken installierten Koppelkomponenten keine oder nur wenige Einstellungsanpassungen.

DiffServ Codepoint Values (DSCP)					DSCP Value <=> Media Type Assignment						Mapping	
Beschreibung	Abkürzung	Binäre Struktur	DSCP		Default <sup>(1)</sup>	Signaling	Voice	Fax <sup>(2)</sup>	Video	IM	802.1D Priority	IP- Precedence (informational)
			Hex	Dez								
Best Effort	be	000-00-0-00	00	00	I					U	0	00
Assured Forwarding Low Drop Prec (Class 4)	af41	100-01-0-00	22	34					E		4	80
Class Selector 5	CS5	101-00-0-00	28	40		E					5	A0
Expedited Forwarding (Per-Hop Behavior)	ef	101-11-0-00	2E	46			E	E			5	A0

Legende:

**ccc-dd-0-uu**

- ccc = class,
- dd = drop precedence,
- 0 = muss auf den Wert = 0 gesetzt werden
- uu = currently unused - muss auf den Wert = 0 gesetzt werden

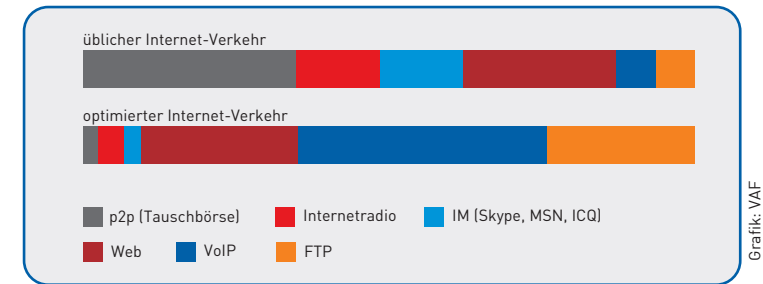
- E= erforderlich
- U= Wertebereich noch nicht endgültig festgelegt
- I= Nur zur Information

- (1) Default-Werte für alle Verkehrstypen, die nicht explizit genannt werden
- (2) Fax via G.711 oder T.38 over UDP; T.38 over TCP werden in der Regel nicht unterstützt

## 4.5 Traffic-Shaping

Oftmals geben Unternehmen einen Großteil Ihres IT-Budgets für teure und geschäftskritische Applikationen aus, um dann herauszufinden, dass diese über WAN und Internet nicht effizient funktionieren. Auch der herkömmliche Internet-Zugang wird unnötigerweise überdimensioniert, nur weil man Bandbreite nicht unterteilt und Dienste nicht priorisiert. Diese Anwendungen müssen sich ihre Bandbreite von weniger wichtigen Applikationen erkämpfen, da die Bandbreiten-Ressourcen beschränkt sind. Die Philosophie der »Freien Fahrt für alle Anwendungen« verzögert wichtige Informationen, behindert die Leistung der Anwendungen und schadet dadurch dem Unternehmen. In der Regel fehlt den Netzen keine zusätzliche Bandbreite, sondern Wege, die vorhandene Bandbreite zu managen und deren Verbrauch zu kontrollieren. Hier setzt das so genannte Traffic-Shaping an.

Das Traffic-Shaping ist ein Telekommunikationsverfahren, bei dem beim Senden der Datenfluss nach definierten Kriterien gesteuert wird. Das Traffic-Shaping arbeitet unidirektional. Im Gegensatz zur Datenflusskontrolle arbeitet dieses Verfahren ohne Steuerinformationen der Gegenseite. In der Regel wird mit Hilfe des Traffic-Shapings der Netzwerkverkehr in verschiedene Bereiche eingeteilt. Diese Einteilungen sind beispielsweise Applikationen oder Protokolle oder Zeit und Art der Verbindung. Das Traffic-Shaping ermöglicht es, gezielt geschäftsrelevante Dienste zu priorisieren. Irrelevante Internet-Dienste können entweder ganz gestoppt oder nur ein geringer Prozentteil der verfügbaren Bandbreite zugewiesen werden.



Exemplarische Darstellung durch Traffic-Shaping optimierter Datenströme

Grafik: VAF

Eine Anwendung oder ein Programm kann über unterschiedliche Ports kommunizieren – von denen es tausend verschiedene gibt. Versucht man nun, Bandbreitenmanagement auf Portbasis zu realisieren, kann das Programm ungehindert über einen anderen freien Port kommunizieren. Moderne Traffic-Shaper erkennen direkt die Kommunikationsströme. Dabei ist es egal, welchen Port das Programm verwendet. Durch die Traffic-Shaper-Technologie wird sichergestellt, dass den geschäftsrelevanten Diensten die höchstmögliche Geschwindigkeit zur Verfügung steht.

Traffic-Shaper nutzen zur Verkehrssteuerung die so genannte Deep-Packet-Inspection, kurz DPI. Auf Basis dieser Filtertechnologie werden die Pakete beziehungsweise deren Inhalte spezifisch nach den Vorgaben des Administrators auf den Schichten 2 bis 7 überwacht beziehungsweise kontrolliert. DPI wird in der Regel als Array von ASIC-Prozessoren realisiert. Diese lassen sich schnell an sich ändernde Bedingungen an-



passen. Dadurch ist der Benutzer in der Lage, bandbreitenintensive und sicherheitskritische Applikationen wie VoIP, P2P, IM und Skype in deren Netzwerken zu kontrollieren. Der Administrator ist in der Lage, anhand der unterschiedlichen Protokolle die vorhandene Bandbreite unter den Anwendungen aufzuteilen. Hierfür lassen sich die jeweiligen Regeln frei miteinander kombinieren.

Durch das Traffic-Shaping werden an den Netzgrenzen bestimmten Anwendungen die notwendigen Bandbreiten zur Verfügung gestellt und andere Applikationen gezielt ausgebremst. Damit werden Lastspitzen ausgeglichen und der Echtzeitverkehr erhält die gewünschte Vorfahrt vor anderen Anwendungen. Da die Switches und Router innerhalb der Netze in der Regel keine echten Traffic-Shaping-Funktionen bereitstellen, arbeiten mit dem Traffic-Shaping in der Regel die DiffServ- oder IntServ- und/oder die Queueing-Funktionen zusammen.

#### ▶ ANMERKUNGEN: TRAFFIC-SHAPING

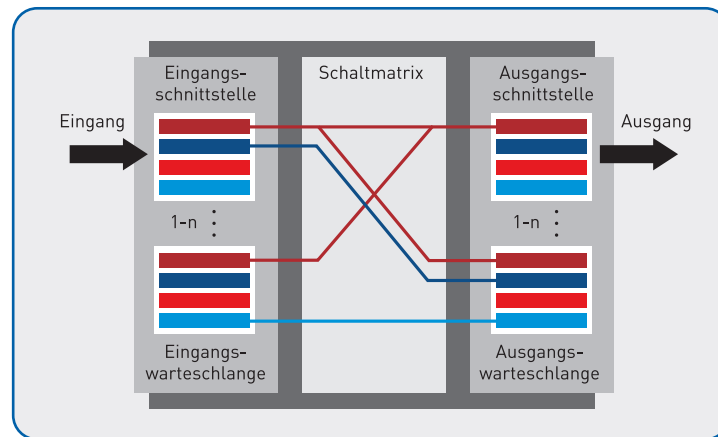
- Das Verfahren vermeidet durch die Bandbreitensteuerung im Edge-Bereich eine Überlastung im Core- und WAN-Bereich und trägt zur (kosten-)optimierten Bereitstellung »richtig dimensionierter« WAN-Strecken bei.
- Es definiert auf Port-Basis die verfügbare Bandbreite pro Applikation. Detailliertes Wissen über den Bandbreitenbedarf einzelner Anwendungen und eine durchgängige Konfiguration der Netzwerkkomponenten sind erforderlich. Wegen proprietärer Elemente sollte das Zusammenspiel der Funktionalitäten von Produkten unterschiedlicher Hersteller geprüft werden.
- Know-how über spezifische Produkteigenschaften, Fehlerortung und Fehlerbehebung ist notwendig.

## 4.6 Queueing

Die Verwendung der Protokollfelder setzt voraus, dass es eine differenzierte Paketbehandlung in Form von Qualitätsklassen in den Netzelementen, den Routern und Switches, gibt. Diese wird durch mehrere Warteschlangen, so genannte Queues, erreicht. Dabei muss jede Warteschlange spezifische Qualitätscharakteristiken, darunter Bandbreite, Verzögerung und Paketverluste, durch die entsprechende Paketbehandlung sicherstellen. Zum besseren Verständnis ist es wichtig, den prinzipiellen Aufbau eines Routers zu verstehen, wie ihn die folgende Abbildung zeigt.

Jeder Router besteht aus einer bestimmten Anzahl an Eingangsschnittstellen und einer bestimmten Anzahl physikalischer Ausgangsschnittstellen. Dazwischen liegt die Schaltmatrix, die anhand der Routing-Entscheidungen die Pakete von einer Eingangsschnittstelle zu einer Ausgangsschnittstelle weiterleitet.

Der Weg der Pakete durch den Router erfolgt in drei Schritten: Nach dem Empfang des Pakets auf der Eingangsschnittstellenkarte wird das IP-Paket in einem Eingangspuffer gespeichert. Dieser Puffer ist notwendig, da die an der Eingangsschnittstelle eintreffende Anzahl der Pakete pro Sekunde größer sein kann als die Rechnerleistung des Routers zur Auswertung des IP-Headers. Die Eingangsdatenrate  $R_{IN}$  steigt, wenn die IP-Pakete wie bei der IP-Telefonie kürzer werden, während die Auswertung des IP-Headers – Abtrennung des Headers, Datenanalyse, IP Headermodifikation – unabhängig von der mittleren Paketlänge ist.



Aufbau eines Routers

Grafik: VAF

Wenn die Auswertung des IP-Headers nicht mehr den herein kommenden Verkehr kurzer Pakete, beispielsweise TCP-Acknowledge mit rund 40 Byte oder VoIP-Pakete mit 50 bis 150 Byte, folgen kann, füllt sich der Eingangsbuffer immer mehr und es kommt beim Überlauf des Pufferspeichers zu Datenverlusten. Auf jeden Fall entsteht durch die Pufferung eine Verzögerung, die unabhängig von Prioritätsklassen ist, da der Router eine Priorisierung erst nach der Auswertung des IP-Headers vornehmen kann. Moderne Router werten den IP-Header nicht mehr in Software, sondern hardwarebasierend in Spezialchips, sogenannten ASICs, in »Wire Rate« aus.

Wire-Rate bedeutet, dass auch sehr kurze Pakete, wie sie bei VoIP vorkommen, ohne Verzögerung ausgewertet und einer Qualitätsklasse zugeordnet werden können. Um verschie-

dene Prioritätsklassen in Routern zu realisieren, besitzen moderne Router eine bestimmte Anzahl an Warteschlangen. Der in der Abbildung oben links dargestellte Router besitzt beispielsweise für jede Eingangsschnittstelle vier verschiedene Warteschlangen, auf die sich vier verschiedene Qualitätsklassen abbilden lassen.

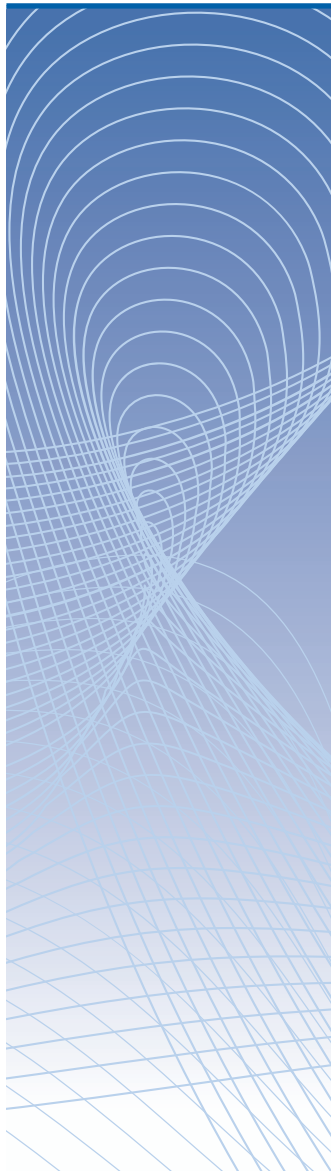
Für jede Warteschlange muss eine QoS-Policy definiert werden. Policies bestehen aus technischen Regeln und Aktionen, um die Einhaltung eines Service-Level-Agreement zu garantieren. In einem Router werden Policies durch eine Paketklassifizierung, Packet-Classifer genannt, einer festgelegten Anzahl von Aktionen und Regeln und einer Steuerung der Warteschlangen, dem Scheduling, ausgeführt.

Die Paketklassifizierung im Router kann Routing-orientiert über die Eingangsschnittstelle oder über die Ursprungs- oder Zieladresse des IP-Pakets erfolgen. Eine weitere Möglichkeit ist die anwendungsorientierte Klassifizierung in Bezug auf den UDP-Ursprungs- oder -Zielport, RTP-Informationen, das DSCP-Feld, die Protokollart oder das MPLS-Label. Eine Zugangskontrolle, Admission-Control genannt, entscheidet, ob eine gewisse QoS-Anforderung zugelassen werden kann.

### Scheduling-Methoden steuern Queueing

Aktionen und Regeln betreffen beispielsweise die Auswahl einer Warteschlange einer bestimmten Länge und Qualitätsklasse, den Schwellwert für das Verwerfen des Pakets, die spezielle Kennzeichnung des Pakets, die DiffServ-Markierung, oder die Bandbreitenlimitierungen, Rate-Limitations genannt. Diese Aktionen und Regeln bestimmen den minimalen, mittleren beziehungsweise maximalen Durchsatz und die Verzögerung





durch den Aufenthalt in der Warteschlange. Eine beispielhafte Router-Konfiguration könnte folgendermaßen lauten: Alle Pakete mit dem Parameter X müssen in der Warteschlange Y gespeichert werden, die eine spezifische Dienstqualität Z (beispielsweise Durchsatz oder Verzögerung) bietet.

Das Problem liegt darin, dass für eine Ende-zu-Ende-Qualität alle Netzelemente verstehen müssen, wie sie Pakete mit dem Parameter X behandeln sollen. Für die Zuteilung der Pakete in die Warteschlangen gibt es die IntServ- und DiffServ-Verfahren. Die Steuerung der einzelnen Warteschlange erfolgt über einen Mechanismus, der entscheidet, welche Pakete welcher Warteschlangen für die Einhaltung der QoS-Anforderungen als nächstes nach einer festgelegten Regel gesendet werden. Dieses Scheduling bestimmt damit das zeitliche Verhalten eines Pakets in einem Router. Für diese Steuerung der Warteschlangen gibt es eine Vielzahl verschiedener Methoden.

▶ **Strict-Priority-Queueing ist eine einfache Methode:** Pakete einer Warteschlange mit geringerer Priorität werden nur dann gesendet, wenn alle Warteschlangen mit höherer Priorität leer sind. Die Warteschlange mit der höchsten Priorität erzielt eine minimale Verzögerung. Diese Methode ist vor allem für VoIP-Verkehr, der möglichst immer Vorrang vor den anderen Daten haben sollte, geeignet. Allerdings kommen die niedrig priorisierten Warteschlangen nie zum Zuge, wenn der hoch priorisierte Verkehr die gesamte Bandbreite benötigt. Daher wird meistens der hoch priorisierte Verkehr auf eine maximale mittlere Bandbreite durch im Anschluss beschriebene Traffic-Shaping-Maßnahmen begrenzt, um auch niederpriorisierten Verkehr zuzulassen.

▶ **Bei Round-Robin unterscheidet man zwei Varianten:** Bitwise-Round-Robin (BRR) versucht, eine kanalorientierte Verbindung zu emulieren, indem jede Warteschlange die Anzahl der empfangenen Byte zählt. Pakete werden beispielsweise nach ihrer Ankunftszeit in Bezug auf die minimale Byte-Länge priorisiert. BRR hat einen schwerwiegenden Nachteil: Wenn eine Warteschlange keine Pakete über einen längeren Zeitraum empfängt, akkumuliert diese die Rechte, die nächsten Pakete als Burst zu senden, und verzögert damit die anderen gleichzeitigen Pakete der anderen Warteschlangen. BRR diskriminiert Anwendungen mit kürzeren Paketlängen. Beim Weighted-Round-Robin (WRR) wird eine Warteschlange laufend nach dem Gewicht geleert, wodurch eine relative Differenzierung der Warteschlangen entsteht. Round-Robin hat den Vorteil, dass es sehr einfach zu implementieren ist.

▶ **Gewichtete Priorisierung** oder **Weighted-Fair-Queueing** weist jeder Warteschlange eine gewisse Bandbreite zu, die bei der Abarbeitung der Warteschlangen dementsprechend berücksichtigt wird. WFQ ist für VoIP nicht gut geeignet, da die »Fairness« sicherstellt, dass nicht priorisierter Verkehr auch behandelt wird. WFQ bietet eine vorhersehbare Verzögerung und ist komplex zu implementieren.

▶ **Class-Based-Queueing** (CBQ) ist ein Schedulingmechanismus, der eine Hierarchie implementiert. Warteschlangen können zu Gruppen kombiniert werden und Gruppen können sowohl relativ zueinander als auch Warteschlangen innerhalb einer Gruppe priorisiert werden.

▶ **Link Fragmentation and Interleaving** (LFI) stellt in Netzen, in denen neben RTP-Sprachdaten auch sehr große Datenpakete übertragen werden, sicher, dass es zu keinem VoIP-Pa-



ketverlust durch Jitterbuffer-Overflow kommen kann. Wenn große Datenpakete, auch mit niedrigerer Priorität, erst einmal in einer Netzkomponente über die Warteschlange in den Sendestatus gelangt sind, dann kann die benötigte Sendedauer die Dimensionierung eines RTP Jitterbuffers überschreiten und es führt zu Paketverlusten im VoIP Gerät.

Da Verkehrsspitzen in TCP/IP-Netzen nicht vermeidbar sind, ist es notwendig, die maximale mittlere Bandbreite zu steuern, um die Policies einhalten zu können. Dies ist vor allem für den Verkehr höchster Priorität notwendig, um einerseits den Verkehr der anderen Qualitätsklassen nicht ganz zu unterbinden und andererseits Störungen des höher priorisierten Verkehrs zu vermeiden. Höhere Protokollschichten wie TCP passen bei Paketverlusten automatisch den Verkehrsfluss an die verfügbare Bandbreite an. Moderne Router bieten eine spezielle Verkehrssteuerung für jede einzelne Warteschlange. Um einen der Policy nicht konformen Verkehr zu erkennen, wird beispielsweise ein Token-Bucket-Mechanismus implementiert. Dieser verwendet die beiden Parameter Größe des Token-Buckets in Bit und mittlere Datenrate in Bit/s. Bei dem Token-Bucket-Mechanismus wird einer Warteschlange jede Sekunde eine gewisse Anzahl an Tokens zugewiesen. Sollte die Warteschlange voll sein, werden allerdings keine Tokens zugewiesen. Jedes Paket in der Warteschlange verbraucht jene Anzahl an Tokens, die der empfangenen Paketlänge entsprechen.

Falls keine Tokens vorhanden sind, wird das Paket verworfen. Ein neuer Token wird alle  $1/r$  Sekunden erzeugt. Zusätzlich dazu gibt es eine Obergrenze  $s$  für die Anzahl der Tokens, um einem Dienst nicht die Möglichkeit zu bieten, zu viele To-

kens für eine zu nutzende Bandbreite zu sammeln. Der Parameter  $s$  repräsentiert dabei die maximale Verkehrsspitze.

Ein sensitiver Ansatz zum Verwurf von Paketen ist Random-Early-Detect. RED verwirft mittels eines dynamischen Mechanismus eine zufällige Anzahl an Paketen, bevor der festgelegte Schwellwert für einen Füllgrad einer Warteschlange überschritten wird. Der Nachteil von RED ist, dass Verstopfung, die durch TCP-Verkehr verursacht wird, nur verschlimmert wird, da der TCP-Mechanismus das Paket wiederholt sendet. Bei UDP funktioniert der Mechanismus hingegen gut.

### Intelligente Steuerung verhindert Verluste

Weighted-Random-Early-Discard (WRED) ist eine Erweiterung von RED, wobei in Warteschlangen höherer Priorität weniger Pakete verworfen werden, als in Warteschlangen geringerer Priorität. Sinnvoll ist es bei TCP, in den Routern und Switches, wenn möglich, den neueren Mechanismus Explicit-Congestion-Notification (ECN) einzusetzen. Dieser Mechanismus verwirft so lange wie möglich keine Daten, sondern setzt im IP-Header ein Bit, welches dem Empfänger signalisiert, dass ein Engpass entsteht. Der Empfänger teilt dann dem Sender über Bits im TCP-Header mit, dass dieser seine Senderate reduzieren soll. Wenn sich alle Verbindungen an diese Regelung halten, kann ein verlustarmes lokales Netzwerk erreicht werden.

Der zweite Schritt auf dem Weg der Pakete durch den Router ist die Weiterleitung des Pakets durch die Schaltmatrix zur Ausgangsschnittstelle. Dazu wird die Ausgangsschnittstelle durch Auswertung des IP-Headers über eine Routing-Tabelle ausgewählt.



Im dritten Schritt wird das IP-Paket auf der Ausgangsschnittstellenkarte in eine Ausgangswarteschlange gespeichert, bevor es zum nächsten Netzelement gesendet wird. Die Ausgangswarteschlange ist nötig, weil nicht ausgeschlossen werden kann, dass die Daten mehrerer Eingangsschnittstellen zu einer Ausgangsschnittstelle weitergeleitet werden müssen.

### Server basierte Call Admission Control (CAC)

Server basierte Call Admission Control (CAC) stellt in Netztopologien mit Bandbreitenengpässen (beispielsweise Anbindung von Zweigstellen über WAN Links) auf Basis einer Funktion des Call Control Servers im Data Center eine geeignete Methode zur Verfügung, um bei mehreren konkurrierenden Verbindungswünschen die Anzahl der Verbindungen dergestalt zu regulieren, dass die erfolgreich aufgebauten Verbindungen eine ausreichende Bandbreite zur Verfügung gestellt bekommen.

Call Admission Control (CAC) liefert eine Bandbreiten-Management Lösung zur Sicherstellung, dass Media Calls nur dann aufgebaut werden, wenn die nötige Bandbreite verfügbar ist. Dadurch werden Verbindungen mit schlechter Qualität vermieden. Somit wird durch CAC ein Mechanismus bereitgestellt, der für Infrastrukturen mit limitierter Bandbreite die Verbindungen im IP-Netzwerk verwaltet und kontrolliert.

Durch ein zentralisiertes Management und Konfiguration erfolgt eine intelligente Bandbreitenreservierung. Hierdurch lassen sich beispielsweise auch Notrufe übermitteln, wenn das erlaubte Bandbreitenlimit bereits erreicht ist.

## 4.7 Sicherung der Dienstgüte im WAN

Das Multiprotocol-Label-Switching, kurz MPLS, ermöglicht die verbindungsorientierte Übertragung von Datenpaketen in einem verbindungslosen Netz entlang eines zuvor aufgebauten Pfades. Dieses Vermittlungsverfahren wird überwiegend von Service-Providern im WAN eingesetzt. Beim MPLS erfolgt die Verlagerung der Routing-Funktion vom Network-Layer in den Data-Link-Layer durch den Einsatz von speziellen Switches. Das Label-Switching (LS) ermöglicht ein Routing/Switching mit Hilfe von Labeln. MPLS bietet den Providern verschiedene Dienste, um den steigenden Datenverkehr zu bewältigen.

- ▶ Vereinfachung und Erhöhung des Durchschaltens von Paketen durch Layer-3-Switching-Technik.
- ▶ MPLS unterstützt Traffic-Engineering und ermöglicht dadurch die Vergabe von Service-Level-Garantien, Service-Level-Agreement genannt.
- ▶ MPLS unterstützt vermaschte Netztopologien.
- ▶ MPLS kann Ressourcen-Reservierungen durchführen.
- ▶ Service-Provider können mit Hilfe von MPLS Serviceklassen anbieten.

Sobald ein IP-Paket ein MPLS-Netz erreicht, wird es am Ingress-Router mit einem zusätzlichen MPLS-Header versehen. Der MPLS-Header wird je nach Technologie innerhalb des bestehenden Headers oder separat eingefügt. Die Zusatzinformation wird als Label bezeichnet. Das Label entspricht einer 20 Bit langen Identifikation. Anhand dieser Identifikation wird eine unidirektionale Verbindung, auch Pfad genannt, eindeutig



bestimmt. Der Wert jedes Labels ist an die Forward-Equivalence-Class (FEC) gebunden, die die unterschiedlichen Übertragungseigenschaften des Übertragungskanals widerspiegelt. Die Anbindung des Labels an eine FEC wird als Label-Binding bezeichnet.

Die Verwendung von MPLS ist nur in MPLS-Domänen oder in Netzwerken möglich, die MPLS transparent übertragen.

## 4.8 Stolpersteine im Bandbreitenmanagement

Multimedia-Anwendungen erfordern einen unterbrechungsfreien, verlustarmen und verzögerungsarmen Transport der Datenströme. Die in den vorigen Kapiteln 4.1 bis 4.7 beschriebenen Mechanismen stellen – mit ihren spezifischen Vor- und Nachteilen – ein breites Repertoire zur Sicherung der Dienstgüte in Netzen bereit. Die folgenden Ausführungen sollen den Praktiker auf häufig übersehene Stolpersteine hinweisen.

### Sicherung der Dienstgüte auf der Ebene 2

Der IEEE 802.1p/Q Standard definiert Funktionen für die Priorisierung von MAC-Paketen auf der Ebene 2. Pakete mit niedrigerer Priorität werden nicht durch das Netzwerk geschwächt, wenn Pakete mit höherer Priorität in den Queues des Netzwerkelementes gepuffert sind. Der Standard beschreibt kein Protokoll zur Zugangskontrolle. Somit ist es möglich, allen Paketen die höchste Priorität zu geben, was allerdings dann zu einer »Verstopfung« des Netzes führen würde.

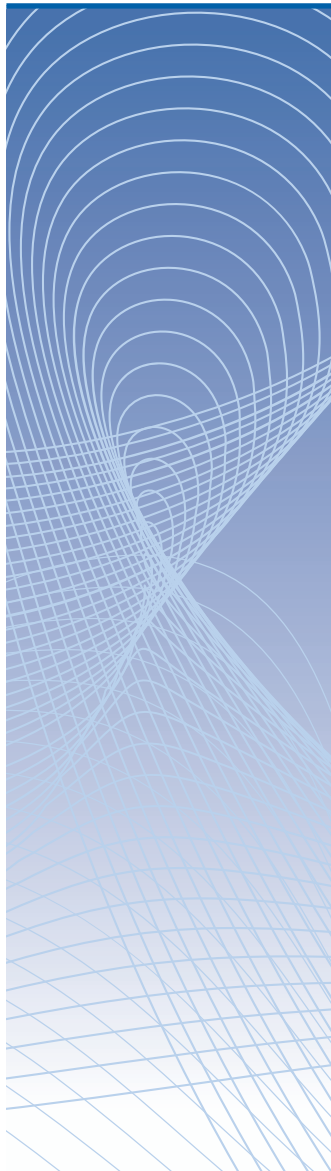
Um eine Ende-zu-Ende-Verbindung mit QoS benutzen zu können, müssen alle Netzwerkelemente im Pfad den Standard

802.1p/Q unterstützen und auch vom Administrator aktiviert und konfiguriert sein.

Der Standard IEEE 802.1p definiert 8 Verkehrsklassen und 7 verschiedene Verkehrsarten. Daraus ergibt sich die benötigte Anzahl von 7 Queues innerhalb der Layer-2-Switches. Derzeit unterstützen die installierten Produkte nach Kenntnis der Autoren acht (ältere Produkte nur zwei oder vier) interne Queues. Dadurch muss unter Umständen ein Mapping der User-Priority in die Verkehrsklassen auf die Anzahl der zur Verfügung stehenden Queues erfolgen. Dadurch ergibt sich zwangsläufig eine Zusammenfassung unterschiedlicher Anforderungen an Dienstgüten innerhalb der verfügbaren Queues, was die Möglichkeiten der individuellen Priorisierung der Verkehrsklassen gegenüber den Möglichkeiten des Standards einschränkt.

Für die Einstellungen der entsprechenden Filter und Queuing-Mechanismen gelten die im Absatz »Sicherung der Dienstgüte auf der Ebene 3« beschriebenen Regeln und Hinweise.

**Achtung:** Die Erweiterung des MAC-Headers um vier Bytes führt in nicht kompatiblen Netzelementen zur Verwerfung der Pakete. Netzelemente, die nicht 802.1p/Q fähig sind, interpretieren die Tagged Frames als fehlerhaft. Darüber hinaus muss die maximale Länge eines Frames auf 1522 Bytes erhöht werden. Lösungen sind der Verzicht auf den Einsatz nicht kompatibler Netzelemente oder die Option des transparenten Weiterleitens der Tagged-Frames.



### Sicherung der Dienstgüte auf der Ebene 3

DiffServ bietet verschiedene Dienste an, um Informationen mit der geforderten Güte zu übertragen. Die Unterscheidung in Assured Forwarding PHB, Expedited Forwarding PHB und Default PHB mit definierten DSCP-Werten ermöglicht eine einfache Implementierung.

In den meisten installierten Netzwerken steht kein Bandwidth-Broker als zentrale Kontrolle zur Verfügung. Dies entspricht dem Gedanken, dass bei der Realisierung von Differentiated Services auf Mechanismen zur Bandbreiten-Reservierung verzichtet werden kann. Die Priorität eines Paketes wird anhand der Header-Information erkannt und dem entsprechenden Per Hop Behaviour zugeordnet.

Die Überwachung und die Vergabe von Bandbreitenressourcen entsprechend den vereinbarten SLAs erfordert, dass sich der Netzadministrator mit folgenden praxisrelevanten Themen eingehend auseinandersetzt:

▶ **Queuing-Mechanismen:** Eine Voraussetzung für die Implementierung der Dienstgüte in Paket-basierten Netzwerken ist eine Veränderung in der Be-/Verarbeitung der Pakete und die Unterscheidung der Pakete oder Datenströme innerhalb eines Netzwerkes. Die Queuing-Verfahren dienen der effizienten Abarbeitung der Pakete oder Datenströme innerhalb aktiver Netzelemente. Die Effizienz bzw. Dienstgüte im laufenden Netzbetrieb wird darum auch durch die vom Netzadministrator gewählten Queuing-Verfahren bestimmt.

▶ **Anzahl der verfügbaren Queues:** Der DiffServ-Standard definiert theoretisch 64 unterschiedliche Verkehrsklassen. In den Standarddokumenten wird jedoch nur von 11 verschiedenen Verkehrsarten gesprochen. Demzufolge ist eine Abbildung al-

ler Verkehrsarten momentan nicht möglich. Aufgrund der verfügbaren Chipsets unterstützen die installierten Produkte nur acht (ältere Produkte nur zwei bis vier) interne Queues. Dadurch muss vom Administrator ein Mapping der Verkehrsklassen auf die Anzahl der zur Verfügung stehenden Queues und ein Zusammenfassen der Datenströme mit unterschiedlichen Anforderungen an die Dienstgüten vorgenommen werden.

▶ **Klassifizierung und Markierung der Pakete:** Die Klassifizierung und das Markieren der Pakete für die nachfolgende QoS-Verarbeitung erfolgt immer am Eingang der Layer-3-Komponente. Die meisten Produkte nutzen hierfür Access Control Lists (ACLs). Das sind Steuerungslisten, die spezifische Funktionen (beispielsweise Paketfilter) auf dem hierfür spezifizierten Port ausführen. Layer-3-Komponenten nutzen zur Klassifizierung und zur Markierung des eingehenden Datenverkehrs so genannte Ingress Filter. Mit Hilfe dieser Filter werden die am Eingang des L3-Switches empfangenen Datenpakete analysiert und die DSCP-Werte entsprechend den vorgegebenen Bandbreitenregeln umgeschrieben.

Die ACLs werden in der Regel hardwarenah abgearbeitet und beeinflussen die Verarbeitungs- bzw. Übermittlungsgeschwindigkeit nicht. Hierbei muss jedoch beachtet werden, dass die hardwarenahe Verarbeitung enge Grenzen hat. Je mehr Filterregeln/Statements ein ACL enthält bzw. je mehr ACLs im jeweiligen Gerät abgelegt werden, desto größer ist die Wahrscheinlichkeit, dass diese nicht mehr in den hierfür vorgesehenen Prozessbereich des ASICs passen. Die Layer-3-Komponente arbeitet dann solche überlangen bzw. zu vielen ACLs in der Software ab. Damit sind einige relevante Konsequenzen verbunden:



- Nur kurze und wenige ACLs werden ohne Geschwindigkeits-einbußen problemlos abgearbeitet.
- Zu lange und zu viele ACLs führen zu drastischen Performance-Einbrüchen und reduzieren die Verarbeitungsgeschwindigkeit der Netze.
- Die Grenzen (Anzahl und Länge) der ACLs sind bei den jeweiligen Produkten von der aktuellen Hardware- und Firmware-Revision abhängig und sind meist in den zugehörigen Release Notes zur Software dokumentiert.
- Bei der Überschreitung dieser »magischen« ACL-Grenzen, erhält der Administrator keinerlei Warnungen vom System. Er wird dies nur anhand der Resultate (fehlende Netzperformance) bemerken.

#### 4.9 Zur Diskussion: DiffServ oder IEEE 802.1p/Q?

Sowohl DiffServ als auch IEEE 802.1p/Q stellen in sich geschlossene Lösungen dar. Während DiffServ auf Ebene 3 des ISO/OSI-Referenzmodells arbeitet, setzt IEEE 802.1p/Q auf der Ebene 2 an. Immer wieder wird die Frage gestellt, welches Verfahren zum Einsatz kommen sollte. Nachfolgend sollen dazu einige Hinweise gegeben werden, ebenso wird aufgezeigt, dass sich heutzutage eine Kombination beider Verfahren nicht immer vermeiden lässt.

Ein Vorteil von IEEE 802.1p/Q ist die klare Zuordnung der Priorität zu den transportierten Informationen. Die Einteilung in die Verkehrsklassen ermöglicht einen direkten Bezug zu den

Anwendungen der Nutzer, z.B. Sprachübertragung via VoIP-Telefon. Das Berücksichtigen von Informationen, die der Administration und Steuerung des Netzwerkes dienen, ist ebenfalls von Vorteil. IEEE 802.1p/Q gestattet dem Netzadministrator somit eine effiziente Kontrolle der Netzbelastungen. Die Einteilung der Informationen in die entsprechenden Serviceklassen ermöglicht die differenzierte Steuerung der Netzwerkbelastung, wie sie in Sprachnetzen z.B. tageszeitabhängig üblich ist.

IEEE 802.1p/Q ist jedoch nicht für die Anwendung in Netzwerken geeignet, die die Größe eines LANs bzw. Segmentes überschreiten. Daher ist das Einsatzgebiet von IEEE 802.1p/Q meist auf lokale Netze und auf Bereiche der Automatisierung (Steuerung, Kontrolle, Messwerterfassung) beschränkt.

DiffServ wird in IP-basierten Netzen angewandt, in denen eine definierte Dienstgüte realisiert werden soll. Das DiffServ-Modell garantiert das Weiterleiten der Pakete unabhängig von deren tatsächlichen Inhalten. Die Entscheidung über die Prioritätsstufe der Pakete obliegt dem Administrator bzw. dem Netzbetreiber. Das DiffServ-Modell stellt dem Betreiber im Hinblick auf die Einführung einer definierten Dienstgüte im Gesamtnetz somit ein flexibleres System zur Verfügung.

Da moderne Netzwerke allerdings vielfach aus einer Kombination von Layer-2-Switches und/oder Layer-3-Switches/Routern bestehen, muss sich der Administrator eines solchen Netzkonstrukts mit den Ende-zu-Ende-Bandbreitenanforderungen auf den unterschiedlichen Ebenen eingehend auseinander setzen und dafür die beiden Verfahren »IEEE 802.1p/Q« und »DiffServ« kombinieren können.

## 5. Quality-of-Experience

Im Mittelpunkt des Quality-of-Experience-Ansatzes steht die vom Nutzer empfundene Qualität der Anwendung. Die Umsetzung subjektiver Kennwerte in technische Messwerte und dynamische Anpassungsregeln soll die Nutzerzufriedenheit erhöhen.

Auch der Begriff der Quality-of-Service, kurz QoS, knüpfte ursprünglich gemäß Definition der Internet-Engineering-Task-Force IETF an die subjektive Wahrnehmung des Nutzers an. Demnach wurde QoS als ein Satz von Attributen verstanden, mit deren Hilfe die Anforderungen von Anwendungen und gleichermaßen deren Nutzern auf die Netzwerke übersetzt werden sollen. Inzwischen hat sich die Definition dahingehend gewandelt, dass die Nutzer unberücksichtigt bleiben und QoS sich nur noch auf die Interaktionen zwischen Anwendungen und Netzwerken bezieht. Dies dürfte vielen technisch orientierten Diskussionsteilnehmern entgegen kommen, kann man doch die – per Definition notwendiger Weise – subjektiven Faktoren der Nutzerzufriedenheit ausklammern. Damit traten die subjektiven Elemente in den Hintergrund der Diskussion und schufen Raum für das Entstehen eines davon abgesetzten Qualitätskonzepts, welches die Subjektivität bewusst ins Zentrum stellt.

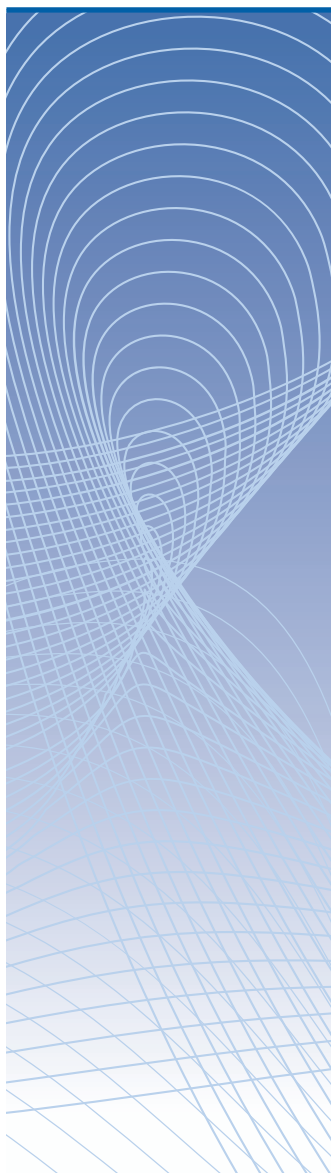
Betrachtet man zunächst den Begriff der Quality-of-Experience, kurz QoE, auch bekannt als Quality-of-User-Experience, dann bezeichnet dieser sehr weit gefasst Maßstäbe und Regeln

eines Anwenders zur Beurteilung und Steuerung von empfangenen Leistungen. Man könnte dies in folgende Fragen übersetzen: »Haben die gelieferten Leistungen die Erwartungen des Nutzers erfüllt?« Ist dies nicht der Fall: »Welche Veränderungen an der Leistungsbereitstellung können vorgenommen werden, um die Erfahrungen des Nutzers zu verbessern?«

QoE ist somit auch im Rahmen von Telekommunikationsnetzen als ein Bewertungs- und Maßnahmenkatalog zu verstehen, mit dem die jeweiligen Dienstleistungen aus der Sicht von Anwendern erfasst werden. Die jeweilige Qualität hängt in der praktischen Umsetzung stark von folgenden Faktoren ab:

- ▶ dem jeweiligen Endgerät des Anwenders,
- ▶ der jeweiligen Umgebungsbedingung,
- ▶ den jeweiligen Ansprüchen und Erwartungen des Benutzers sowie
- ▶ der Art des Inhalts und seiner Bedeutung.

Ziel des gemischt objektiv-subjektiven Bewertungsverfahrens ist es, mehrere Metriken zu generieren, um zukünftig dem Anwender die Leistungen so gut wie möglich bereitstellen zu können.



nen. Den Begriff Quality-of-Experience hat Microsoft im Zusammenhang mit dem Produkt »Office Communications Server« (OCS) in die Kommunikationstechnik eingeführt. Zugleich ist einerseits zu berücksichtigen, dass über Lizenzmodelle OCS und der QoE-Ansatz inzwischen auch Eingang in Produkte anderer Hersteller gefunden haben. Und andererseits darf die nachfolgende Darstellung nicht so verstanden werden, als wenn jeder der vorgestellten Mechanismen nur im OCS realisiert würde. (Zu den QoE-orientierten Techniken, die auch von anderen Herstellern implementiert werden, gehören: Wideband Audio (7 kHz), verschiedene Funktionen im Media Stack wie: Acoustic Echo Cancellation (AEC), Dynamischer Jitterbuffer, Packet Loss Concealment (PLC).) Da nach Kenntnis der Autoren jedoch ein umfassender QoE-Ansatz derzeit nur von Microsoft verfolgt wird, beziehen sich die Aussagen in dieser Abhandlung auf OCS (Version R2) und basieren wesentlich auf technischen Informationen von Microsoft.

QoE ist in der Umsetzung von Microsoft ein Lösungsansatz, der primär auf Endgeräte ausgerichtet ist. Das eigentliche Transportnetz ist von diesen Mechanismen nur nachrangig betroffen. Die Grundlage für das Funktionieren von Quality-of-Experience bilden das Monitoring der Echtzeitdatenströme und deren dynamische Verbesserung, um dem Nutzer die aktuell im Regime bestmögliche Qualität zur Verfügung zu stellen. Zu den wesentlichen Eckpunkten gehören:

▶ **Benutzerorientierter Ansatz der »empfundenen Qualität«:** Die von den Nutzern subjektiv empfundene Qualität der empfangenen Echtzeitströme steht im Mittelpunkt der Lösung und umfasst alle relevanten Parameter zur Verbesserung der Datenströme.

▶ **Intelligente Endpunkte:** So genannte »smart adaptive« Endpunkte analysieren die jeweiligen Übertragungsbedingungen und passen die hierfür notwendigen Parameter an. Hierbei unterscheidet Microsoft zwischen einem Endpunkt (Gerät – beispielsweise einem PC -, auf dem ein Communicator Client lauffähig ist) und Endgeräten (das Ende des Kabels oder der Wireless Strecke - beispielsweise das Headset). Der QoE-Ansatz funktioniert nur dann durchgängig, wenn man alle an einer Kommunikation beteiligten Teile kennt. Der Communicator erkennt welches Endgerät angeschlossen ist und kann anhand dieser Information seine Berechnungen/Annahmen machen

▶ **Permanente Kontrolle der Benutzerqualität:** Kontinuierliche Messungen, eine permanente Quantifizierung und das Monitoring der gesamten Benutzerdaten soll zur Verbesserung der empfundenen Qualität beitragen.

QoE bricht mit dem traditionellen, netzorientierten QoS-Lösungsansatz. Das Netzwerk ist nicht mehr alleiniger »Herrscher« über die Übermittlungsparameter, denn in den QoE-Lösungsansatz fließt ein erweitertes Spektrum der für die Ende-zu-Ende-Übertragung notwendigen Parameter ein: Netzwerkparameter, Hardware, Applikationen, psychologische und physikalische Aspekte. Die damit verbundenen Effekte werden grundlegend in zwei Bereiche unterteilt:

▶ **Netzwerkeffekte:** Diese wirken sich über die gesamte Übertragungsstrecke Ende-zu-Ende aus. Auf diesen Bereich hat sich die Verbesserung der Übertragungsqualität bisher konzentriert.

▶ **Payload-Effekte:** Payload-Effekte haben ihre Ursachen in den spezifischen Inhalten der Datenpakete. Auf diesen Bereich haben beispielsweise die Umweltbedingungen und die Qualität der Endgeräte einen erheblichen Einfluss. Hierzu zählen





beispielsweise Echo, Übersprechen und Nebengeräusche. Dieser Bereich wurde bei dem bisherigen netzwerkzentrierten QoS-Modell ignoriert.

## 5.1 Breitbandige Sprache zur Verbesserung der Telefonlösungen

Die QoE-Lösung bricht auch mit vielen Basisparametern der regulären Telefonie und versucht neue Wege zu gehen. Beispielsweise überträgt eine Microsoft-UC-Umgebung durch die Default-Einstellung die Telefonströme immer als breitbandige Sprache. Im Gegensatz dazu wird in traditionellen Telefonnetzen und VoIP-Techniken die Sprache immer schmalbandig übermittelt.

In klassischen Telefonnetzen wird die Sprache mit einer Frequenz von 8 kHz abgetastet. Hieraus ergibt sich ein nutzbarer Frequenzbereich von 4000 Hz. Nach Abzug des Schutzbands ergeben sich die Bereiche zwischen 200 Hz bis 3400 Hz als nutzbares »Telefonie-Band«. Die so komprimierte Sprache ist ausreichend für eine ordentliche Verständigung der Kommunikationspartner, geht jedoch einige Kompromisse im Bereich der subjektiven Sprachqualität ein. Beispielsweise bewegt sich die Basisfrequenz bei typischen Männern und Frauen unter 200 Hz. Erwachsene Männer nutzen eine Basisfrequenz zwischen 85 Hz und 155 Hz und erwachsene Frauen zwischen 165 Hz und 255 Hz. Diese Basisfrequenzbereiche werden jedoch nicht mehr von einem schmalbandigen Übertragungssystem übermittelt. In diesen Übertragungssystemen wird dieser Mangel durch die Übermittlung von genügend Oberschwingungen kompensiert

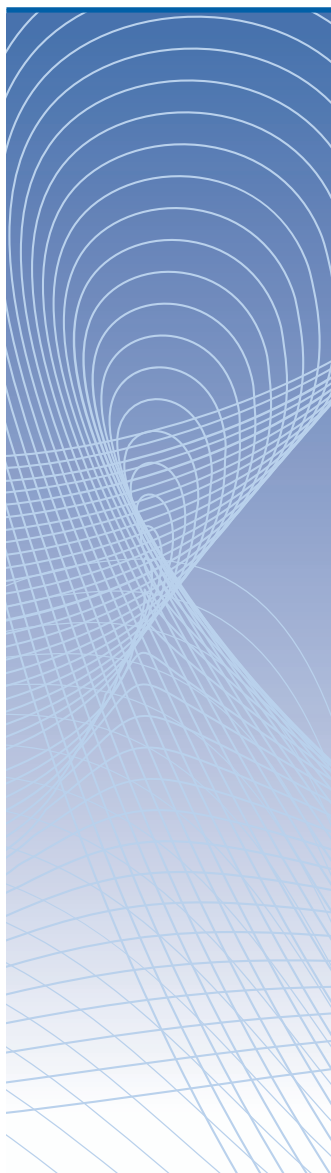
und das eigentliche Stimmbild wird ungefähr reproduziert.

Die »RTAudio«-Breitband-Codecs von Microsoft nutzen eine Abtastfrequenz von 16 kHz. Daraus ergibt sich ein Frequenzbereich von 8000 Hz und ein nutzbarer Frequenzbereich von rund 50 Hz bis 7000 Hz. Dadurch ist eine präzisere Übertragung der niedrigen und hohen Frequenzbereiche möglich. Dies kommt der Qualität des Sprachsignals zugute. Eine breitbandigere Übermittlung der Sprache verbessert nicht nur die Verständlichkeit und Natürlichkeit der Sprache, sondern gibt dem Zuhörer auch ein Gefühl der Transparenz.

## 5.2 Intelligente, adaptive Endpunkte

Bestimmte Netzfehler und die daraus resultierenden Verfälschungen des übertragenen Signals lassen sich mit Hilfe von Tools für das Netzwerkmanagement in den Griff bekommen. Die Beseitigung der Übertragungsfehler bedarf jedoch der entsprechenden Werkzeuge und Mechanismen auf der Applikationsschicht. Aus diesem Grund müssen die Endpunkte einer QoE-VoIP-Lösung autonom die Ende-zu-Ende-Qualität kontrollieren und gegebenenfalls korrigierend eingreifen. Die Endgeräte in traditionellen VoIP-Lösungen sind in der Regel sehr einfache Geräte und bieten nur wenige On-Board-Fähigkeiten. Typischerweise verfügen diese Geräte nur über begrenzte Rechenleistung und Speicher.

Die Microsoft-UC-Lösung basiert auf dem Microsoft-Office-Communicator-2007 und zusätzlichen Anwendungen, die auf intelligenten Endpunkten arbeiten. Diese Endpunkte – in der Regel PCs – verfügen meist über genügend Rechenleistung und Speicher zur Ausführung zusätzlicher Anwendungen. Den Kern



der Plattform bildet der Mediastack des Office-Communicators. Dieser verfügt über die folgenden Merkmale.

▶ **Authentifizierte und verschlüsselte Kommunikationskanäle:**

Die Authentifizierung und Verschlüsselung der Signalisierung und der eigentlichen Nutzdatenströme gehört zu den Grundfunktionen der UC-Lösung. Zusammen mit weiteren Funktionen, beispielsweise der Unterstützung der Interactive-Connectivity-Establishment-Spezifikation (ICE), einem Mechanismus zur Überwindung von NAT-Gateways, ist der Zugriff auf die Systeme von jedem beliebigen Ort aus möglich.

▶ **Permanente Quantifizierung der QoE:** Das System quantifiziert und überwacht kontinuierlich die tatsächlichen QoE-Werte. Hierzu werden die von den Endgeräten erzeugten Messdaten ausgewertet.

▶ Der Mediastack unterstützt eine Reihe zusätzlicher Codecs und Funktionen, die die Auswirkungen der netz- und nicht-netzspezifischen Störungen ausgleichen.

### 5.3 Messungen und Monitoring

Der Office-Communicator misst die aktuellen Übertragungsparameter und generiert daraus die individuelle Metrik für die Verbindungen. Der Office-Communications-Server sammelt und aggregiert diese in den Call-Detail-Records (CDR) als Metrics-CDR. Die Metrics-CDRs umfassen verschiedene Parameter.

▶ **MOS-Werte:** Die Effekte des Netzwerks als auch der Payload werden auf Basis spezieller Algorithmen ermittelt. Diese Werte werden beispielsweise in den Bewertungsfaktor MOS als Qualitätskriterium der Sprachqualität umgerechnet. Die Kom-

munikationsendpunkte überwachen somit kontinuierlich die subjektive und objektive Benutzerqualität. Darüber hinaus werden auch entsprechende Reports für alle Endpunkte und Verbindungen erstellt, die als Qualitätsmerkmale zur Trendermittlung genutzt werden können.

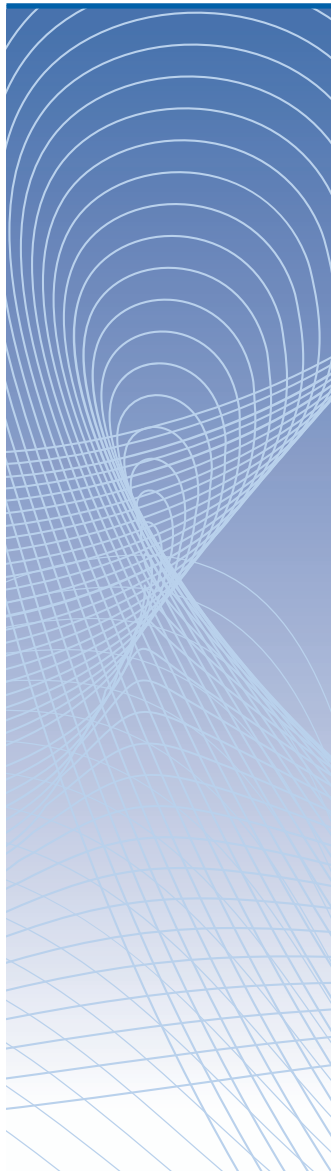
▶ **Network-Service-Quality-Parameter:** Das System generiert Messwerte von allen relevanten NSQ-Parametern. Dadurch erhält der Netzadministrator exakte Kennzahlen über die verfügbare Netzperformance.

Der OCS Monitoring Server sammelt die Metrics-Call-Detail-Records aller Systeme und analysiert die Daten. Auf Basis der ausgewerteten Daten werden dem Systemadministrator die Ursachen für Probleme aufgezeigt und entsprechende Alarme generiert. Auf Überlastungen wird reagiert, indem der Mediastack alle Verbindungen dynamisch an die neue Situation anpasst. Dies kann beispielsweise durch eine geringe Reduzierung der Sprachqualität aller Sprachverbindungen geschehen.

### 5.4 Mediastack

Der UC-Mediastack enthält Funktionen zur Verbesserung der Übertragungsqualität und zur angepassten Ausschöpfung der vorhandenen Netzwerkressourcen.

▶ **Real-Time-Audio-Codec (RTAudio):** Der Breitband-Sprachcodec wird in der Default-Einstellung der Microsoft-Plattform immer genutzt. Darüber hinaus werden alle bisher in der VoIP-Welt üblichen Codecs ebenfalls unterstützt. Der Encoder lässt sich entweder im Schmalband- mit 8 kHz Sampling-Rate oder im Breitband-Betrieb mit 16 kHz Sampling-Rate einsetzen. Da-



rüber wird der Codec mit einer konstanten Bit-Rate (CBR) oder mit einer variablen Bit-Rate (VBR) genutzt. Im VBR-Modus ergibt sich eine verbesserte Effizienz, da sich der Codec- und der Mediastack dynamisch auf die übermittelten Sprachsignale anpassen.

► **RTAudio-Encoder:** RTAudio ist ein Subband-Encoder. Die Zahl der Unterbänder ist abhängig von der Abtastfrequenz. Für eine Abtastfrequenz von 8 kHz wird nur ein einziges Band genutzt. Bei Abtastfrequenzen > 8 kHz verwendet der Codec mehrere Bänder. In der Regel werden die meisten Informationen bei der Sprachübermittlung im unteren Band übertragen. Aus diesem Grund wird die Mehrzahl der Informationen im unteren Übertragungsband gesendet. Der Encoder ermöglicht darüber hinaus auch eine dynamische Verwaltung der Breite der jeweiligen Bänder. Bei Bedarf wird die Bandbreite im unteren Band erhöht, während die nutzbare Bandbreite im höheren Übertragungsband schrittweise sinkt.

Das Eingangssignal wird mit Hilfe von Filtern in Subband-Signale aufgeteilt. Das Rate-Control-Modul bestimmt die Kodierung für die Subbänder. Diese hängt von mehreren Faktoren ab: der Signalcharakteristik der einzelnen Subbänder, den verfügbaren Bitstream-Puffern und der zu erzielenden Ziel-Bit-Rate. Im Allgemeinen werden weniger Bit für die Codierung »einfacher« Frames, also stimmlose und leise Passagen, und entsprechend mehr Bit für »komplexe« Frames benötigt.

Der Encoder für jedes Subband basiert auf mehreren Codebuchblöcken, einem Linear-Prediction-Coefficients-Analyseblock (LPC) und einem Synthese-Filter. Es sind im System mehrere vordefinierte Arbeitsbereiche festgelegt. Diese werden

durch die Kombination der verschiedenen Codebücher für die unterschiedlichen Sampling-Raten bestimmt. Das Rate-Control-Modul bestimmt den Verarbeitungsmodus für jeden Frame. Der Encoder enthält auch eine optionale Funktionseinheit zur Integration von Fehlerkorrekturmechanismen sowie zur Integration redundanter Informationen in den RTAudio-Bitstream. Arbeitet der Codec im Forward-Error-Correction-Modus, werden zusätzliche Informationen in den Bitstream eingefügt. Der RTAudio-Decoder arbeitet in einem Pull-Modus. Das Jitter-Control-Modul ermöglicht aktives Management des Paket-Jitters und der Paketverluste.

### Automatische Fehlerunterdrückung inklusive

Der Decoder besitzt auch Fähigkeiten zur Unterdrückung von Fehlern. Die Fehlerunterdrückung in Kombination mit einer Jitter-Kontrolle trägt zu einer Verbesserung der Audio-Qualität bei. Die aktuelle Bit-Rate des Multi-Rate-Codec wird durch den Qualitätskontroller und die dynamische Anpassung an die tatsächlichen Übertragungsbedingungen bestimmt. RTAudio nutzt sechs Breitband- und drei Schmalband-Übermittlungs-raten. Die Standard-Bit-Rate im Breitband-Modus beträgt 45 kBit/s. Die mittlere Bit-Rate im Breitband-Modus beträgt zwischen 24 kBit/s und 45 kBit/s. Wird die FEC-Redundanzfunktion genutzt, beträgt die mittlere Bit-Rate im Breitband-Modus zwischen 43 kBit/s und 74 kBit/s. Die Standard-Bit-Rate im Schmalband-Modus beträgt 28 kBit/s. Die mittlere Bit-Rate im Schmalband-Modus beträgt zwischen 15 kBit/s und 28 kBit/s. Wird die Redundanzfunktion genutzt, erhöht sich die mittlere Bit-Rate im Schmalband-Modus auf 25 kBit/s bis 40 kBit/s.



RTAudio arbeitet effizienter als herkömmliche Codecs wie G.711, G.729 oder G.726 und benötigt bei gleicher Audioqualität deutlich weniger Bandbreite.

Codec	Paketgröße (ms)	Gesamte Payload (Bit/s)	Bit-Rate ohne Redundanz (Bit/s)
RTAudio (16 kHz)	20	29000	45000
RTAudio (8 kHz)	20	11800	27800
G.711	20	64000	80000
G.729	20	32000	48000
G.726	20	8000	24000

Vergleich der Bit-Raten traditioneller Codecs

(Quelle: Microsoft)

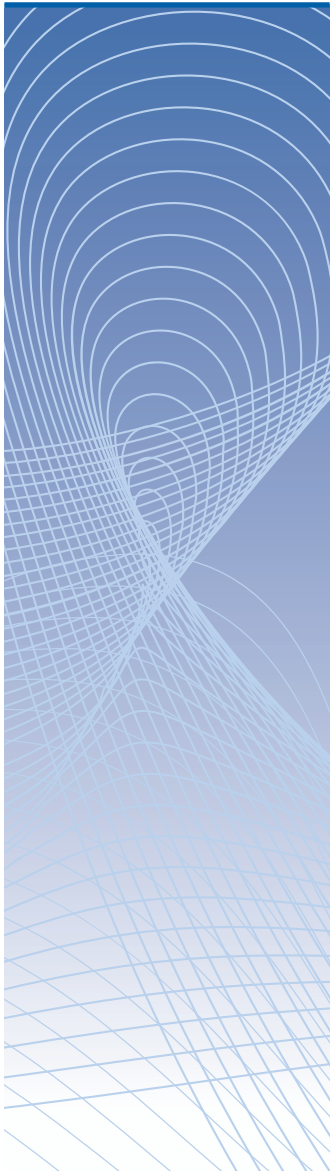
Der Quality-Controller (QC) ist eine Software-Komponente, die sich auf Basis der von den unterschiedlichen Schichten des Mediastacks zur Verfügung gestellten Informationen dynamisch an das Verhalten des Absenders anpasst. Der QC berücksichtigt eine Vielzahl von Parametern. Dazu gehören Anzahl und Art der Datenströme, die aktuelle Schätzung der verfügbaren Bandbreite, die verfügbaren Codecs und die NSQ-Bedingungen. Zu den Optimierungsfunktionen gehört auch die Forward-Error-Correction (FEC).

Die QC wählt während des Gesprächs dynamisch die genutzten Codecs aus und passt sich in Echtzeit an sich ändernde Übertragungsverhältnisse an. Bei der temporären Überlastung eines Links wird der SRC in jedem Endpunkt der Session aktiv und erkennt, dass sich die Netzbedingungen verschlechtert

haben. Jeder Mediastack verringert daraufhin schrittweise die Bit-Rate. Zwangsläufig wird dadurch auch die Qualität der Audiodaten schrittweise gesenkt. Da die Microsoft-Codecs gegenüber den Standard-Codecs eine höhere Qualität liefern, führt eine Reduzierung der Bit-Raten nicht zu einer drastisch verschlechterten Sprachqualität. In der Regel werden auch dann noch bessere Qualitätsergebnisse erzielt, wenn die Standard-G.729-Codecs auf Grund fehlender Übertragungsbandbreite nicht mehr ordnungsgemäß arbeiten würden.

Die Automatic-Gain-Control (AGC) erkennt die Sprachenergie des Eingangssignals und passt die analogen/digitalen Signale an, um ein komfortables Hören für den Benutzer zu ermöglichen. Forward-Error-Correction (FEC): Unter Kontrolle der QC sorgt der RTAudio-Encoder für einen zusätzlichen Schutz vor Datenverlust während der Übertragung. Hierzu werden redundante Informationen in die RTAudio-Bitstreams integriert. Anhand dieser Zusatzinformationen ist der Empfänger in der Lage, während der Übertragung verloren gegangene Datenpakete zu rekonstruieren. Durch die FEC-Funktion steigt die für die Übertragung notwendige Datenrate natürlicherweise an. Die Endpunkte der Verbindung entscheiden deshalb dynamisch über die Nutzung der FEC-Funktion für den jeweiligen Datenstrom.

Die Acoustic-Echo-Cancellation (AEC) ist ein adaptiver Filter zur Reduzierung von Rückkoppelungen zwischen Lautsprecher und Mikrofon. Der Acoustic-Echo-Cancellation-Filter verhindert, dass eventuell auftretende Echos den Empfänger erreichen. Auf der Empfängerseite wird die Empfangsqualität durch verschiedene Module verbessert:



- ▶ **Redundanzfunktion:** Rekonstruktion des Signals auf Basis der FEC-Funktionen.
- ▶ **Dynamische Jitter-Puffer:** Die Jitter-Puffer passen sich dynamisch an die Audiogeschwindigkeit des Ausgangs an. Die Dynamik des Prozesses beseitigt automatisch die Auswirkungen der Zwischenpufferung bei Datenströmen mit geringem Jitter und sorgt für reibungslose Übergänge, wenn sich die Jitter-Bedingungen ändern. Dies ermöglichen variabel lange Jitter-Puffer.
- ▶ **Packet-Loss-Concealment (PLC):** Dieses Modul bereitet die empfangenen Sprachsignale so auf, dass verloren gegangene Pakete bei geringen bis mittleren Verlustraten zu keiner oder nur zu einer geringfügigen Verschlechterung der Ausgangsqualität führen.

## 6. Fazit

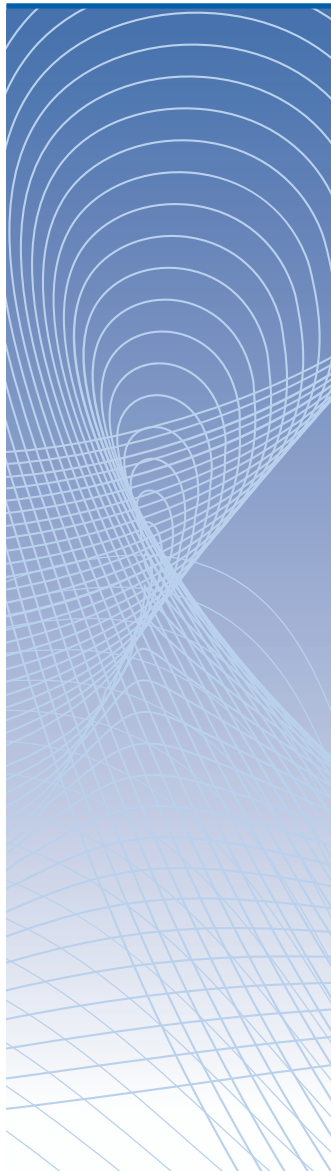
Die in diesem Dokument vorgestellten Verfahren des Bandbreitenmanagements sorgen für eine Verbesserung der Sprachqualität und Dienstverfügbarkeit bei der Übermittlung von VoIP-Datenströmen.

Die Verfahren unterscheiden sich teilweise fundamental beziehungsweise bedingen die Kombination einzelner Teillösungen. So macht beispielsweise der Einsatz von DiffServ ohne Trafficshaping im WAN keinen Sinn. QoE ist ein in die Endgeräte integrierender Lösungsansatz. Dieser setzt keine durchgängigen QoS-Mechanismen im Netzwerk voraus, sondern nutzt diese, so sie vorhanden sind. Keine Technik bietet für sich genommen ein Allheilmittel und befriedigende Lösungen ergeben sich grundsätzlich aus der richtigen Auswahl und Implementierung von Verfahren für die jeweilige Netzsituation, deren Überwachung und Nachjustierung im laufenden Betrieb sowie vorausschauenden Anpassungen bei Erweiterungen oder Strukturänderungen des Netzes, bei der Implementierung neuer Dienste und Anwendungen.

- ▶ **Ausreichend verfügbare Netzbandbreite ist obligatorisch:** Weder QoE noch QoS sind ein Ersatz für zu wenig verfügbare Übertragungskapazitäten im Netz. Eine ausreichend dimensionierte Bandbreite im Netzwerk trägt zur Reduzierung der Verzögerungen bei und verringert die Paketverluste.
- ▶ **Richtige Provisionierung ist eine Optimierungsaufgabe:** Auch großzügig bemessene Bandbreiten im LAN/WAN bieten keine Sicherheit für die Einhaltung der erforderlichen beziehungs-

weise definierten Dienstqualität. Erratisch auftretende Lastspitzen sind nahezu unvermeidbar. Für sie verantwortlich sind beispielsweise neue Applikationen, Software-Updates, Funktionsstörungen in Koppelkomponenten oder Änderungen des Nutzerverhaltens. Auch ist es eine Binsenweisheit, dass der IP-Traffic kontinuierlich anwächst. Die zunehmende Etablierung von Multimedia-Anwendungen verstärkt diese Entwicklung zusätzlich. Letztlich stellt die Bereitstellung von Bandbreiten(-reserven) einen iterativen Optimierungsprozess dar, in dem ein »Zuviel« an vorsorglichen Reserven die Betriebskosten unnötig, weil ohne Gewährleistung des gewünschten Effekts der Qualitätssicherung belasten. Speziell im WAN und an den Grenzen zu den öffentlichen Netzen hat die Dimensionierung unmittelbare Effekte auf die laufenden Betriebskosten und deren kalkulatorische Derivate wie RoI oder TCO. Der Einsatz von QoS-Techniken ist darum ein geeignetes Instrument für die effiziente Ausnutzung der jeweils gegebenen Bandbreite bei gleichzeitiger Sicherung der Dienstqualität.

- ▶ **DiffServ hat die Marktrelevanz von IntServ überholt:** Immer noch wird der aus der klassischen Telekommunikation kommende IntServ-Ansatz als »eigentlich überlegener« QoS-Mechanismus in Diskussionen hochgehalten. Allerdings bietet der



Großteil der aktuellen Produktwelt und darauf ausgerichtete QoS/CoE-Lösungen keine Unterstützung für IntServ-Mechanismen. DiffServ-Mechanismen haben sich in der Praxis als kostengünstiger, einfacher zu konfigurieren und zu verwalten erwiesen, sind skalierbarer und reichen als Grundlage für ein umfassendes Bandbreitenmanagement im LAN und WAN aus. Auf der Ebene 3 sollte darum nur DiffServ eingesetzt werden.

- ▶ **Elastizität in die Lösung einbauen:** Der Sättigung beziehungsweise Überlast einer oder mehrerer Verbindungen im Übertragungsweg muss mit entsprechenden Queueing-Verfahren in den Switches/Routern begegnet werden. Dadurch wird gewährleistet, dass die hoch priorisierten Verkehrsströme auch bei Überlastung der Transportressourcen übermittelt werden.
- ▶ **Proprietäre Implikationen berücksichtigen und abwägen:** Filterbasiertes Traffic-Shaping trägt effektiv zur Verkehrsflusssteuerung bei. In Ermangelung von Standards setzt dabei jeder Hersteller eigene Filterregeln um. In der Auswahl der Komponenten und insbesondere in der Implementierungspha-

se muss darum auf eine stimmige Konfiguration aller Koppelkomponenten geachtet werden. In dem Microsoft-OCS basierenden CoE-Ansatz verhindern RTAudio, FEC und das dynamische Auswechseln der Codecs, dass bei knapper Bandbreite netzbedingte Überlastungen auftreten. Dadurch wird den Ursachen von Jitter, Paketverlusten und Verzögerungen dynamisch begegnet. Bei diesen Erweiterungen muss sich der Netzbetreiber jedoch im Klaren darüber sein, dass die CoE-Mechanismen nur auf Microsoft-lizenzierten Produkten zur Verfügung stehen.

- ▶ **Fundament der Qualitätssicherung sind verlässliche, objektive Informationen:** Die Gewährleistung von Verfügbarkeit des Dienstes Sprache in definierter Qualität erfordert Messungen in der Implementierungsphase und beim Monitoring im laufenden Betrieb. Nicht zuletzt sollte auch berücksichtigt werden, dass es durchaus Unterschiede zwischen Datenblatt und zugehöriger Netzwerkkomponente geben kann. Während in der Vergangenheit Voice-orientierte Vormessungen und Monitoring

Mechanismus	Wirksamkeit	Standard	Erfordert Ergänzungen	Skalierung	Verfügbarkeit	Anwendbarkeit WAN
Overprovisioning	Layer 1 bzw. Layer2	802.3	-	gut	ja	problematisch
802.1p/Q	Layer 2	802.1p/Q	-	mittel	ja	problematisch
DiffServ	Layer 3	RFC 2474, RFC 2475	-	gut	ja	gut
IntServ	Layer 3	RFC 2210	RSVP	gut	nein	sehr gut
Traffic Shaping	Layer 2 bis Layer 7	proprietär	DiffServ	gut	Herstellerspez. Varianten	sehr gut
CoE	Layer 5 bis Layer 7	proprietär	DiffServ	gut	nur OCS	gut

Tabelle: Vergleich der QoS-Technologien



auf Grund hoher Kosten fast ausschließlich bei Großinstallationen und im Carrier-Umfeld als Bestandteil der Überwachung von SLAs durchgeführt wurden, bietet der Markt heute auch ein Spektrum mittlerer bis klein dimensionierter Lösungen und darauf aufbauender Dienstleistungen. Daraus sind wichtige Impulse für eine deutliche Verbesserung der Qualitätssicherung in kleinen und mittelständischen Firmennetzen zu erwarten.

### Schlussbemerkung

Historisch waren die Netze der Sprachkommunikation (PSTN, PBX) und der Datenkommunikation (Ethernet, Internet) getrennt. Die Sprachnetze auf Basis von Time-Division-Multiplexing (TDM) bildeten separierte, robuste Infrastrukturen, die für die Erfordernisse der Echtzeitkommunikation optimiert waren und sich durch hohe Dienstverfügbarkeit und ausgezeichnete Sprachqualität als Marktstandard auszeichneten. Die Datenetze waren durch hohe Entwicklungsgeschwindigkeit, Flexibilität und höhere Offenheit der Systeme charakterisiert. Zugleich wiesen sie aber auch eine größere Fehleranfälligkeit auf und eigneten sich nur sehr bedingt für den Transport echtzeitkritischer Dienste. Die heute viel zitierte Konvergenz von

IT und TK (ITK-Konvergenz) ist präziser beschrieben als die technologische Integration der Telekommunikation in die TCP/IP-basierte Netzwelt. Aus den Netzen für Datenkommunikation werden umfassende Kommunikationsnetze, welche die zeitkritischen Anforderungen der Sprachkommunikation mit bedienen sollen, woraus sich erheblicher Nachholbedarf ergibt. Mit Verfahren des Bandbreitenmanagements lassen sich dafür Lösungen schaffen, die auch für den geschäftlichen Anwender in Firmennetzen funktionieren. Dies ist ein wichtiger Entwicklungsschritt und die Voraussetzung, um weitere Effizienzsteigerungen technisch unterstützter Kommunikationsprozesse zu erschließen.



## Literatur und Weblinks

802.1Q-2005

**IEEE standard for local and metropolitan area networks virtual bridged local area networks**

[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?tp=&isnumber=34308&arnumber=1637340&pnumber=10905](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&isnumber=34308&arnumber=1637340&pnumber=10905)

802.1D-2004

**IEEE Standard for Local and metropolitan area networks Media Access Control (MAC) Bridges**

[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?tp=&isnumber=29062&arnumber=1309630&pnumber=9155](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&isnumber=29062&arnumber=1309630&pnumber=9155)

*O. Aboul-Magd, S. Rabie*

**RFC 4115: A Differentiated Service Two-Rate, Three-Color Marker with Efficient Handling of in-Profile Traffic**

<http://www.ietf.org/rfc/rfc4115.txt?number=4115>

*J. Babiarz, K. Chan, F. Baker*

**RFC 4594: Configuration Guidelines for DiffServ Service Classes**

<http://www.ietf.org/rfc/rfc4594.txt?number=4594>

*Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, E. Felstaine*

**RFC 2998: A Framework for Integrated Services Operation over DiffServ Networks**

<http://www.ietf.org/rfc/rfc2998.txt?number=2998>

*S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss*

**RFC 2475: An Architecture for Differentiated Services**

<http://www.ietf.org/rfc/rfc2475.txt?number=2475>

*R. Bless, K. Wehrle*

**RFC 3754: IP Multicast in Differentiated Services (DS) Networks**

<http://www.ietf.org/rfc/rfc3754.txt?number=3754>

*R. Braden, D. Clark, S. Shenker*

**RFC 1633: Integrated Services in the Internet - Architecture and Overview**

<http://www.ietf.org/rfc/rfc1633.txt?number=1633>

*R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin*

**RFC 2205: Resource ReSerVation Protocol (RSVP) —Version 1 Functional Specification;**

<http://www.ietf.org/rfc/rfc2205.txt?number=2205>

*K. Chan, J. Babiarz, F. Baker*

**RFC 5127: Aggregation of DiffServ Service Classes**

<http://www.ietf.org/rfc/rfc5127.txt?number=5127>

*Cisco Systems*

**Comparing Traffic Policing and Traffic Shaping for Bandwidth Limiting**

[http://www.cisco.com/en/US/tech/tk543/tk545/technologies\\_tech\\_note09186a00800a3a25.shtml](http://www.cisco.com/en/US/tech/tk543/tk545/technologies_tech_note09186a00800a3a25.shtml)

*Cisco Systems*

**Quality of Service Networking**

<http://www.cisco.com/en/US/docs/internetworking/technology/handbook/QoS.pdf>

*M. Eder, H. Chaskar, S. Nag*

**RFC 3387: Considerations from the Service Management Research Group (SMRG) on Quality of Service (QoS) in the IP Network**

<http://www.ietf.org/rfc/rfc3387.txt?number=3387>

*F. Le Faucheur*

**RFC 4124: Protocol Extensions for Support of DiffServ-aware**

MPLS Traffic Engineering

<http://www.ietf.org/rfc/rfc4124.txt?number=4124>

*Jörg Fischer*

**VoIP-Praxisleitfaden - IP-Kommunikation für Sprache, Daten und Video planen, implementieren und betreiben**

Hanser Fachbuch, München, erste Auflage, 2008

*D. Grossman*

**RFC 3260: New Terminology and Clarifications for DiffServ**

<http://www.ietf.org/rfc/rfc3260.txt?number=3260>

*Mathias Hein*

**TCP/IP Protokolle**

MITP, Bonn, sechste Auflage, 2004

*J. Heinanen, F. Baker, W. Weiss, J. Wroclawski*

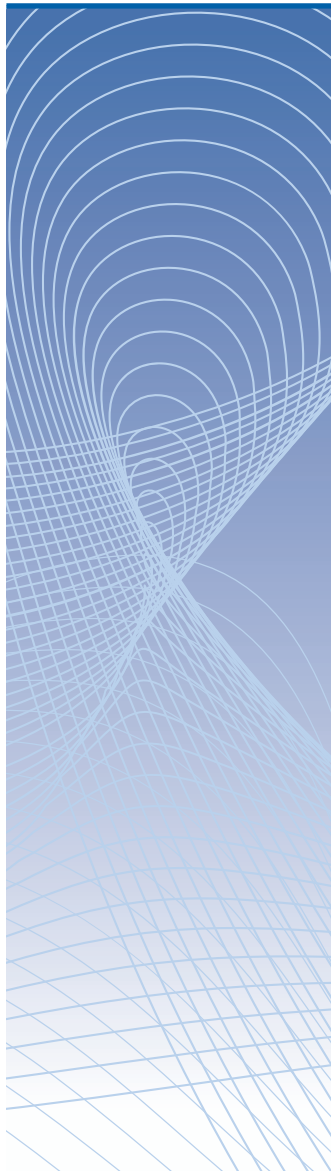
**RFC 2597: Assured Forwarding PHB Group;**

<http://www.ietf.org/rfc/rfc2597.txt?number=2597>

*G. Huston*

**RFC 2990: Next Steps for the IP QoS Architecture**

<http://www.ietf.org/rfc/rfc2990.txt?number=2990>



*V. Jacobson, K. Nichols, K. Poduri*  
**RFC 2598: An Expedited Forwarding PHB**  
<http://www.ietf.org/rfc/rfc2598.txt?number=2598>

*Benjamin Kolbe*  
**Ein neues parametrisiertes Modell für QoS bei VoIP**  
 Diplomarbeit Fachhochschule Flensburg, 08.10.2008

*Microsoft*  
**Quality of Experience, A strategic competitive advantage of Microsoft Unified Communications**  
 Microsoft Corporation 2007  
[http://download.microsoft.com/download/3/a/3/3a367b0d-4fff-44c5-9125-3ac5c037f195/OCS\\_QoE.doc](http://download.microsoft.com/download/3/a/3/3a367b0d-4fff-44c5-9125-3ac5c037f195/OCS_QoE.doc)

*B. Moore, D. Durham, J. Strassner, A. Westerinen, W. Weiss*  
**RFC 3670: Information Model for Describing Network Device QoS Datapath Mechanisms**  
<http://www.ietf.org/rfc/rfc3670.txt?number=3670>

*K. Nichols, B. Carpenter*  
**RFC 3086: Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification**  
<http://www.ietf.org/rfc/rfc3086.txt?number=3086>

*T. Rottenau, B. Stütz*  
**Schneller schalten - 10-Gigabit-Ethernet-Triple-Play-Switches**  
 Network Computing 22 - 4 (2007)

*T. Rottenau, B. Stütz*  
**Schneller schalten - 10-Gigabit-Ethernet-Triple-Play-Switches, Teil 2**  
 Network Computing 30 - 11-12 (2007)

*T. Rottenau, B. Stütz*  
**Wettlauf im LAN - Vergleichstest Gigabit-Ethernet-Switches**  
 Network Computing 18 - 3 (2008)

*Toralf Runge*  
**Unterstützung von Quality of Service in Ethernet-Netzwerken - Evaluierung der Mechanismen und Ableitung von Konfigurationsempfehlungen**  
 Diplomarbeit, Fachhochschule Stralsund, 2003

*Stefan Schulze*  
**Mechanismen zur Sicherung der Dienstgüte und deren Bedeutung in den Netzwerken der nächsten Generation**  
 Diplomarbeit, Fachhochschule Stralsund, 2003

*Siemens Enterprise Communications*  
**Quality of Service (QoS) Default Setting Profile**  
 Siemens Enterprise Communications GmbH & Co KG, München  
 White Paper, Status 03/2010

*Wikipedia*  
**Quality of experience**  
[http://en.wikipedia.org/wiki/Quality\\_of\\_Experience](http://en.wikipedia.org/wiki/Quality_of_Experience)

*Wikipedia*  
**Quality of Service**  
[http://de.wikipedia.org/wiki/Quality\\_of\\_Service](http://de.wikipedia.org/wiki/Quality_of_Service)

*Wikipedia*  
**Traffic-Shaping**  
<http://de.wikipedia.org/wiki/Traffic-Shaping>