

Dienstgarantien für mobile Systeme

Daniel Jankovic, Peter Ruppel

Hauptseminar "Dienste & Infrastrukturen mobiler Systeme"
Wintersemester 03/04
Institut für Informatik
Ludwig Maximilians Universität München
{jankovic, ruppel}@informatik.uni-muenchen.de

Abstract. Diese Seminararbeit behandelt den Begriff der Dienstgüte (Quality of Service, QoS) und stellt gängige Mechanismen und Modelle zur Realisierung von Dienstgarantien vor. Anschließend wird auf die Probleme im Zusammenhang mit der Übertragung dieser Modelle auf mobile (d.h. leiterungebundene) Systeme eingegangen.

In einem zweiten Teil (Kapitel 5 und 6) werden die aktuellen Vorschläge der IEEE 802.11e Arbeitsgruppe sowie das Universal Mobile Telecommunications System (UMTS) in Bezug auf die Bereitstellung von QoS behandelt. Die Enhanced Distributed Coordination Function (EDCF) und Hybrid Coordination Function (HCF) von 802.11e werden vorgestellt und mit den bisherigen Funktionen in 802.11 verglichen. Abschließend wird die QoS-Architektur von UMTS (nach dem 3rd Generation Partnership Project, 3GPP) erläutert.

1 Einleitung

Schon aus den 50er Jahren sind erste Versuche einer Kommunikation zwischen den ersten Rechnern bekannt. In Forschungseinrichtungen, Großbanken oder Fluggesellschaften wurden einzelne Arbeitsstationen mit leistungsfähigen Großrechnern verbunden. Mit der Entwicklung des World Wide Web (www) veränderte sich ebenfalls der Charakter der übertragenen Daten. Das Hauptziel bei der Entwicklung von Rechnernetzen ist die optimale Auslastung eines Netzes, wie auch die Einbeziehung neuer Dienste. Netze müssen in der Lage sein, die gleichzeitige Übertragung von verschiedenen Datenformaten wie Audio, Video und Bilderdaten über das selbe Medium durchführen zu können. Dies beinhaltet die Übertragung zeitkritischer Daten in Echtzeit, sowie die optimale Auslastung der zur Verfügung stehenden Bandbreite bei gleichzeitiger Zuteilung gesicherter Datenraten an die einzelnen Anwendungen. Diese Forderung widerspricht jedoch dem Prinzip, nach dem das Internet bis jetzt arbeitet. Das "Best-Effort-Prinzip" verträgt sich nicht mit den Anforderungen, die zeitkritische Daten stellen. Sie müssen bevorzugt übermittelt werden, damit es nicht zu Verzögerungen und damit Qualitätseinbußen kommt. Dadurch wird die Möglichkeit, alle Daten und Systeme gleich behandeln zu lassen, unterdrückt. Mit der Fortentwicklung von leistungsfähigen Technologien rückte die Bedeutung von QoS in den Brennpunkt

des Interesses. Denn aufkommende zeitkritische Dienste funktionieren in paketorientierten Netzen nur dann, wenn diese Daten Vorrang vor weniger zeitkritischen Informationen erhalten. Der Haken dabei ist was eigentlich unter "Dienstgüte" zu verstehen ist.

Die Idee des Dienstgütevertrages ist entstanden. Die Quelle spezifiziert den generierten Verkehr und verspricht, sich daran zu halten. Das Netz verspricht die Übertragung mit garantierten Dienstgütemerkmalen.

2 Begriffsbestimmung: Quality of Service (QoS)

Unter dem Begriff Dienstgüte (Quality of Service, QoS) werden Eigenschaften eines Dienstes bezeichnet, die ein Teilnehmer fühlen oder messen kann. Zu beachten sind folgende Punkte:

1. QoS wird betrachtet aus der Sicht des Anwenders, welcher den verlangten Dienst Ende zu Ende nutzen möchte und diesen wirklich nutzt. Die Eigenschaften eines einzelnen Netzelements steuern zur Dienstgüte bei, können aber nicht gesondert als QoS - Parameter betrachtet werden.
2. QoS ist auf messbare technische Aspekte begrenzt. Alles was nicht in diesen Rahmen fällt wird nicht betrachtet.
3. In Abhängigkeit von der Zielsetzung kann ein Teil der QoS Parameter betont und getrennt geprüft werden. Es ist nicht immer notwendig, dass alle Parameter berücksichtigt werden.
4. Der Begriff QoS lässt unterschiedliche Definitionen zu und es sind mehrere konkurrierende Begriffsbestimmungen entstanden.

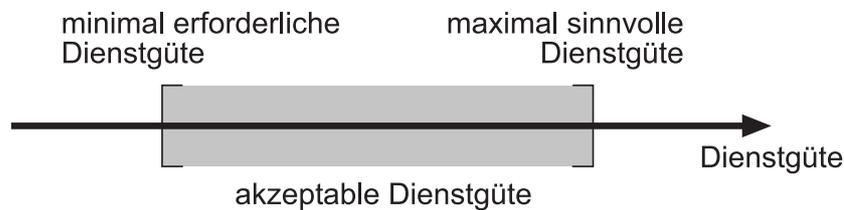


Fig. 1. Dienstgüte

Z.B. definiert [1]: "Dienstgüte kennzeichnet das definierte, kontrollierbare Verhalten eines Systems bezüglich messbarer Parameter".

Cisco bezeichnet QoS "als Fähigkeit eines Netzes, Dienste für bestimmte Anwendungen von einem Ende der Kommunikation zum anderen zur Verfügung zu stellen" [2].

QoS ist im Grunde eine Beschreibung eines bestimmten Dienstes, welcher einem Benutzer angeboten und zugesichert wird, wobei der Benutzer sowohl eine

Person als auch eine Protokollschicht sein kann. Die Eigenschaft, eine Verbindung zu messen, zu verbessern und soweit wie möglich vorhersagen zu können, ist die Bestimmung von QoS.

Der angebotene Dienst kann ein Dienst auf oberster Schicht sein, wie z.B. bei Videokonferenzenanwendungen, oder ein Dienst der unteren Schichten, wie z.B. das verbindungsorientierte TCP/IP-Protokoll. Im Zusammenhang mit QoS werden verschieden Parameter und damit verbundene Fachtermini erwähnt:

1. Datendurchsatz (Throughput): Performanz einer Verbindung anhand der Menge der übertragenen Daten pro Sekunde. (z.B. Pakete/Sekunde bzw. Bytes/Sekunde) Es gilt zu beachten, dass Wartezeiten auf die Quittierung den Durchsatz vermindern.
2. Verzögerung (Delay)
 - Paketverzögerung: Gesamtverzögerung bei der Übertragung vom Sender bis zum Empfänger (Latency).
 - Dabei werden zwei Arten von Verzögerung festgelegt:
 - real delay: Verzögerung aufgrund der begrenzten Ausbreitungsgeschwindigkeit des Lichts oder der Elektronen. Es stellt eine untere Grenze für den Delay dar.
 - induced delay: Verzögerung der Nachricht aufgrund von Bearbeitungszeit in Zwischen- und Endsystemen.
3. Jitter: Variation der Differenz der Übertragungszeit zweier benachbarter Datenpakete. Werden zwei Datenpakete direkt nacheinander vom Sender zu unterschiedlichen Zeitpunkten gesendet, so werden danach diese beim Empfänger entsprechend nacheinander empfangen. Jitter ist bei paketorientierten Netzen nicht zu vermeiden und muss daher beachtet werden. Man unterscheidet zwischen Asynchronen-, Synchronen- und Isochronen Verhalten.
4. Loss: (Verlust)
 - Paketverlustrate : Das Verhältnis der verworfenen Pakete zur Gesamtanzahl der gesendeten Datenpakete (Packet Loss Rate)
 - Sensitivity classes: ignore, indicate, correct losses
 - Verlustrate: maximale Anzahl von Verlusten pro Zeitintervall
 - Verlustgröße: maximale Anzahl von konsekutiv verlorenen Paketen
5. Fairness (Gerechtigkeit): Im bisherigen Datenverkehr wurde das Best-Effort-Prinzip angewandt, welches alle Anwender ebenbürtig behandelt. Mit Einführung von vorhersagbarer Dienstgüte wird diese Grundidee verworfen. Daten hoher Priorität werden gegenüber Daten niedrigerer Priorität favorisiert. Da in Zukunft eine entsprechende Tarifierung dieser Dienste zu erwarten ist, kann man nur noch von einer kostenbasierten Fairness sprechen. Man achtet darauf, dass eine Klassifikation der Ressourcen zwischen zeitkritischen und zeitunkritischen Anwendungen einer gewissen Fairness entspricht.

Bislang gibt es keine übergreifende Definition von "Quality of Service", die für alle Netze und Anwendungsfelder gültig ist. Die Dienstgüte wird durch die Summe der Übertragungseigenschaften des (virtuellen) Kanals bestimmt,

welcher für die Anwendung bereitgestellt wird. Je nach Art des dienst anbietenden Netzes wird zwischen zwei Arten des QoS unterschieden:

2.1 Timeliness QoS

Timeliness QoS beschreibt die Dienstgüte in Abhängigkeit der Übertragungszeit. Aus der Sicht der drahtgebundenen Kommunikation, wie sie z.B. innerhalb eines LANs vorliegt, ist die Zeitabhängigkeit der wesentliche Gesichtspunkt. Im Gegensatz zu Datenkommunikation über Funk stellt hier die reine Verfügbarkeit eines Kanals zur Kommunikation meist kein Problem dar. Das wichtigste Kriterium für die Dienstgüte ist, ob ein verfügbarer Kanal für eine Anwendung dienlich ist.

Der wichtigste Parameter ist hierbei die Höhe des Datendurchsatzes, welcher durch die der Verbindung zugewiesenen Bandbreite bestimmt ist.

Die Latenzzeit (Latency) ist ein Maß für die absolute Verzögerung einzelner autonomer Dateneinheiten zwischen dem Sende- und Empfangsvorgang. Sie gibt also an, wie lange ein Bit benötigt, um vom Sender zum Empfänger zu gelangen. Im Gegensatz hierzu wird durch den Jitter die zulässige Schwankung der Laufzeit beschrieben. Er ist durch unterschiedliche zeitliche Abstände von Dateneinheiten innerhalb eines Datenstroms zu erkennen.

Pakete innerhalb eines Datenstroms können zueinander unterschiedliche Abstände haben. Datendurchsatz und Jitter sind damit für einen isochronen, kontinuierlichen Datenfluss, wie z.B. für den Verkehr eines Videosignals, die entscheidenden Qualitätsparameter. Die Latenzzeit spielt dagegen in Echtzeitkontrollsystemen eine Rolle (z.B. bei der Überwachung verteilter Produktionsabläufe). Dort ist es wichtig, dass eine Steuerinformation innerhalb einer definierten Maximalzeit übertragen wird, d.h. dass sie bei einer Maschine ankommt.

2.2 Connectivity QoS

Connectivity QoS beschreibt die Dienstgüte im Hinblick darauf, wann welche Übertragungskanäle zur Verfügung stehen und Stationen auf das Übertragungsmedium zugreifen können. Die Koppelung der Station an das Übertragungsmedium und die damit verbundene Verfügbarkeit von nutzbaren Kanälen wird im Connectivity QoS erfasst. Die Verfügbarkeit ist vor allem für mobile Dateneinrichtungen wichtig. Es wird daher zwischen Netzen unterschieden, deren Dateneinrichtungen aufgrund ihrer Funktionalität keiner, einer zeitweisen oder einer ständigen Koppelung an das Medium bedürfen.

3 Dienstklassen (Class of Service, CoS)

Der Begriff der Dienstklasse ist im Allgemeinen an eine spezifische Implementierung eines Übertragungssystems gebunden, an dieser Stelle soll er aber implementierungsunabhängig betrachtet werden.

Jede Dienstklasse umfasst jeweils eine Untergruppe erzielbarer QoS-Parameter. Die Vollständigkeit aller Dienstklassen reflektiert die Kondition des Übertragungssystems wieder. Class of Service (CoS) teilt verschiedene Arten von Verkehr in unterschiedliche Klassen (Daten, Sprache, Audio, Video) ein, die bei der Übertragung unterschiedlich behandelt werden. [1] hat drei CoS-Klassen definiert:

1. Best-Effort: Dienste ohne Garantie.
2. Vorhersagbare Dienste: Dienste, welche, indem sie das vergangene Verhalten auswerten, Grenzwerte für zukünftige Dienste festzulegen.
3. Garantierte Dienste: Dienste mit garantierten Leistungen, welche durch Dienstgüteparameter sogenannte Grenzwerte entweder deterministisch oder im statischen Mittel beschrieben werden.

Die Internet Engineering Task Force (IETF) hat zwei QoS Ansätze ausgearbeitet, die sich der Transportschicht zuordnen lassen. Integrated Services und Differentiated Services, beide verfolgen unterschiedliche Wege, um eine Dienstgüte im Netz umzusetzen.

3.1 Integrated Services (IntServ)

Folgender Gedanke liegt dem Integrated Services Modell zugrunde: die IP-Protokollfamilie wird entsprechend integriert, damit unterschiedliche neue Dienste mit einer annehmbaren Güte angeboten werden können. Dafür wird für jeden Datenstrom ein einzelner virtueller Kanal (VC) geschaltet. Ausgehend von den Anforderungen des Datenverkehrs wird ein QoS-Profil mit fest vorgegebenen Verkehrsparametern vereinbart. Dementsprechend werden dann entlang der Route Ressourcen in jeder Übertragungsrichtung reserviert.

3.1.1 Hilfsmittel Um reservierte Dienstgüte zu unterhalten, anzufordern und aufzuheben ist ein Zusammenspiel von mehreren QoS-spezifischen Software- und Hardwarekomponenten erforderlich. Man kann sie als einzelne Bausteine der Integrated Services betrachten. Hier werden diese Bausteine aufgelistet und ihre Aufgaben beschrieben.

3.1.1.1 Token Bucket Filter Ein Token Bucket Filter reguliert die maximal zulässige Datenübertragungsrate. Von einem Tokengenerator werden dabei mit einer konstanten Rate Tokens erzeugt, damit wird ein "Eimer" (Bucket) fortlaufend gefüllt. Wenn ein Datenpaket eintrifft, muss die Anzahl der Token im Bucket mindestens so groß wie das Paket sein. Trifft dieser Fall zu, so wird das Paket als filterkonform erachtet und weiterverarbeitet. Der Inhalt des Buckets

wird um die Größe des Datenpakets reduziert. Umgekehrt wird das Paket als nicht filterkonform behandelt. Stehen keine Datenpakete zum Versenden an, so füllt sich der Bucket. Ist der Bucket voll, so werden neuerzeugte Token verworfen (s. Fig. 2).

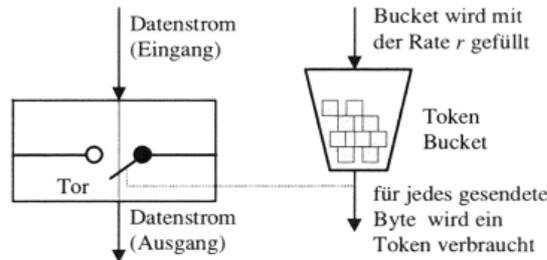


Fig. 2. Token Bucket Filter

Angewandt auf die Datenübertragung bedeutet jetzt r die mittlere Übertragungsrates der Daten und die Größe des Buckets die maximale Burstgröße. Solche Filter lassen sich zur Kontrolle des Datenverkehrs einsetzen. Durch Veränderung der Größe des Buckets oder der Tokengenerationsrate kann der Datenverkehr anhand seiner Häufungen im Datenstrom (Burst) oder anhand der Datenrate gefiltert werden. Man kann durch Reihenschaltung komplexere Filter konstruieren. Außerdem ist es üblich, in Servicespezifikationen den Datenfluss mit Hilfe von Token Bucket Parametern darzustellen.

3.1.1.2 Admission Control = AC Jedes IntServ Endgerät sowie IntServ-fähiger Router muss eine AC Instanz besitzen. Trifft eine Reservierungsanfrage auf ein bestimmtes Interface, so untersucht die AC Einheit, ob genügend Ressourcen zur Verfügung stehen und ob der geforderte Dienst mit dem auf dem zugehörigen Interface eingesetzten Paket Scheduling Verfahren zugesichert werden kann.

3.1.1.3 Policy Control = PC Diese Einheit prüft die Ansprüche der Benutzer in Bezug mit QoS Anforderungen. Außerdem kümmert sie sich um die Authentisierung der Teilnehmer im Netz. Im Bereich der PC herrscht bislang noch eine Unklarheit bezüglich der einzusetzenden Verfahren. Dabei ist ohne Implementierung der PC Einheiten kein ernsthafter Einsatz der Integrated Services denkbar. Die Policy Entscheidungen können zur Zeit anhand der Sender- bzw. Empfänger- IP Adresse und/ oder der jeweiligen Port- Nummer und nicht personenbezogen getroffen werden.

3.1.1.4 Resource reSerVation Protocol = RSVP Das Resource Reservation Protocol wurde 1993 als Reservierungsprotokoll für das Internet entwickelt und

bekam 1997 den Status eines "Request For Comment" (RFC) und wird verwendet für die Steuerung und Reservierung von Bandbreite. Dieses Protokoll ist im RFC 2205 spezifiziert.

RSVP ist empfangenorientiert und kennt Ströme nur in einer Richtung. Daher werden Ressourcen nur für den Datenstrom vom Sender zum Empfänger reserviert. Sollen auch Ressourcen für den Rückkanal reserviert werden, muss der Sender diese selbst anfordern. Ressourcen werden in jedem Knoten reserviert, der auf dem Weg zwischen Empfänger und Sender liegt. Bei einer Multicast-Umgebung werden Anforderungen beim ersten Knoten gesammelt und kombiniert, eine entsprechende neue Anforderung wird weitergeschickt. Dadurch steigt der Overhead logarithmisch und nicht linear. Die beiden wichtigsten Nachrichten sind Path, welches vom Sender geschickt wird, und Resv (Reservation Request), welche den selben weg zurück zum Sender als Bestätigung läuft. Empfängt ein Knoten eine Reservierungsanforderung, so übergibt der dort laufende RSVP-Prozess die Nachricht zunächst an die Admission Control und an die Policy Control. Sind beide Ergebnisse zufriedenstellend, ordnet der Packet Classifier jedes empfangene Datenpaket einer QoS-Klasse zu (Fig. 3). Der Packet Scheduler bestimmt die Abfolge, in der die Datenpakete übergeben werden und trägt somit für die Verwirklichung des QoS die Verantwortung. Anschließend wird eine neue Resv-Nachricht an den nächsten Upstream-Knoten auf dem Pfad zum Sender weitergeleitet.

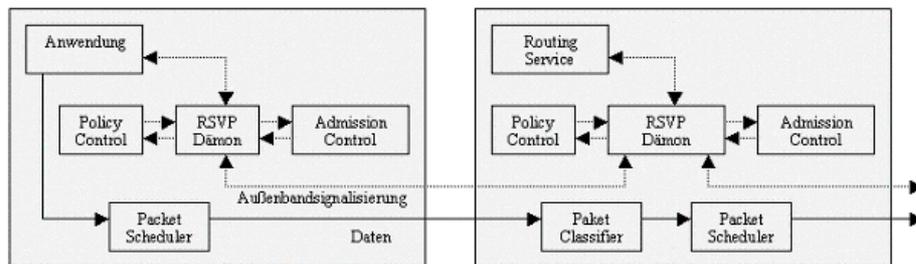


Fig. 3. RSVP

3.1.2 Dienstklassen der ISW Group Von der Integrated Services Working Group der IETF (Internet Engineering Task Force) wurden seit 1997 fünf Dienstklassen (service classes) definiert:

- Controlled Load QoS
- Controlled Delay QoS
- Predictive Delay QoS
- Committed Rate QoS

- Guaranteed QoS

Die wichtigsten von Seiten der Implementierung sind dabei der Controlled Load QoS und der Guaranteed QoS, deswegen gehe ich im Folgenden näher darauf ein.

3.1.2.1 Controlled Load Network Element Service(CL): dieser Dienst "täuscht" dem Nutzer ein unbelastetes Netz in Zeiten der Überlast im Netz vor. Er stellt also eine eher tolerante Variante dar und wird oftmals in der Literatur als "besserer als Best-Effort-Service" bezeichnet. Er ist im RFC 2211 spezifiziert worden. CL-Service kann z.B. für Audio- bzw. Video-Streaming oder für Web-basierte Transaktionen eingesetzt werden. CL ist wegen der Absicht auf einfache Integration in das bestehende Netz nur auf die notwendigsten Funktionen beschränkt.

Durch die Kontrolle der Übertragungsraten, welche von Anwendungen angefordert werden, ermöglicht der Control Load QoS die dedizierte Zuteilung von Bandbreite. Die übrigen QoS-Parameter bleiben unberücksichtigt. Da die maximale Transferrate eines Übertragungsmediums eine konstante Größe ist, ist eine Implementierung des Control Load QoS auch oberhalb der Sicherungsschicht durchführbar. Dies hat zum Vorteil, dass eine weltweite Einführung des Control Load QoS auf Basis von RSVP ohne Änderungen der unteren Schichten und damit zum Großteil ohne Austausch von Hardware erreicht werden kann. Auch wenn allein der Parameter des Throughputs berücksichtigt wird, so nimmt der Control Load QoS damit doch eine Sonderstellung im Vergleich zu den übrigen Serviceklassen ein.

Integrated Services fordert von der Quelle eine Flussspezifikation (flowspec), die in Traffic und Request Spezifikation aufgeteilt ist.

- Traffic Spezifikation (TSpec): beschreibt die Variation der angeforderten Bandbreite über der Zeit. Zur Beschreibung der TSpec wird das Konzept des Token Bucket Filter verwendet.
- 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.

3.1.2.2 Guaranteed Service : dieser Dienst stellt einem Nutzer harte und jederzeit eingehaltene Garantie für Bandbreite und Verzögerung dar. Alle nötigen Parameter werden bevor oder spätestens während Netzdaten übertragen werden ausgehandelt. Dieser Dienst ist im RFC 2212 spezifiziert. Vorstellbare Anwendungen sind Videokonferenzen oder Voice over IP (VoIP). Ein Guaranteed Quality of Service soll einen sehr hochwertigen Dienst bereitstellen.

Unterschieden wird zwischen folgenden Serviceklassen:

- Constant Bit Rate (CBR): Bei dieser Serviceklasse wird eine feste Transferrate (Bitrate) mit maximal zulässige Jitter- und Latency-Werten einer Anwendung zugewilligt. Anspruchsvolle Dienste wie z.B. Videokonferenzen sind erreichbar

- Variable Bitrate (VBR): Die Serviceklasse wird durch die Angaben der mittleren Bitrate sowie einer ggf. maximal auftretenden Burstrate beschrieben. In einer zweiten Ausprägung, dem Realtime VBR (rt VBR), können zudem Angaben für die Latency und den Jitter spezifiziert werden.
- Unspecified Bit Rate (UBR): In dieser Serviceklasse ist es nicht möglich, Vorgaben zu machen. Die maximale Bitrate wird zwar spezifiziert, welche jedoch nicht vom Netz garantiert wird. Sie dient allein der Kalkulation der noch verfügbaren Ressourcen.
- Available Bit Rate (ABR): Ihr stehen nur noch freie Ressourcen zur Verfügung, die nicht durch andere Serviceklassen belegt sind. Im Vergleich zu UBR-Verkehr wird eine minimale Datenrate spezifiziert. ABR-Verkehr genügt daher höheren Anforderungen als UBR-Verkehr.

Es werden bestimmte Anforderungen an den GS-Dienst gestellt:

1. Garantierte Bandbreite.
2. überschreitet der Datenfluss die in der TSpec vereinbarten Grenzwerte nicht, so verhält sich das Netz wie folgt:
 - (a) Es treten keine Warteschlangenverluste in den einzelnen Netzelementen auf. Im Rahmen der Integrated Services werden Verluste in Folge von Fehlern in Netzelementen (Änderung der Route, Ausfall eines Routers, etc.) nicht berücksichtigt. Es wird auch nicht versucht Verluste zu beheben, welche infolge solcher Fehler entstanden sind.
 - (b) Die maximale Verzögerung und die verfügbare Bandbreite sind stabil, was bedeutet dass sie sich nicht ändern solange der Datenpfad erhalten bleibt.
 - (c) Die ausgehandelten Grenzwerte werden nicht von der Abweichung des Datenflusses vom Flussmodell überschreitet.

Unterschiede welche auch der GS-Dienst nicht beachtet:

1. Jitter werden weder berechnet noch berücksichtigt.
2. Nur die maximale Verzögerung und keine mittlere oder minimale Verzögerung werden berechnet. Jedoch wird die maximale Verzögerung immer für die ungünstigsten Verhältnisse mit Reserve berechnet, deswegen werden die meisten Pakete schneller am Ziel ankommen. Dieser Faktor ist bei der Programmierung zeitkritischer Anwendungen bedeutend, die auf Guaranteed Service aufsetzen und ist somit zu berücksichtigen. Daten, welche vorzeitig ankommen, müssen entsprechend verarbeitet bzw. zwischengepuffert werden.

Der Datenfluss wird mit Hilfe des Token-Bucket-Filters beschrieben. Jedes Netzelement längs des Datenpfades innerhalb des Services berechnet unterschiedliche Parameter auf dem entsprechenden Teilabschnitt. Aus diesen Parametern für die Teilabschnitte werden die Ende-zu-Ende QoS-Parameter bestimmt. Am wichtigsten ist die maximale Ende-zu-Ende-Verzögerung.

Die Grundidee des IntServ ist die Reservierung einer festen Bandbreite auf den Übertragungsstrecken für einzelne Teilnehmer.

Die benötigten Ressourcen werden verringert durch prädektive Schätzung des ankommenden Verkehrs. Dies ermöglicht die Annahme, dass die angemeldete Datenrate nicht die ganze Zeit ausgeschöpft wird. Dadurch wird eine bessere Granularität geboten, da die Bandbreite in 1Byte-Schritten variiert werden kann und nicht in 64kbit/s bzw. 2Mbit/s.

Auch ein hoher Grad an Flexibilität wird in IntServ geboten, denn die Ressourcenreservierung kann während der Datenübertragung angefordert, geändert oder aufgelöst werden.

Andererseits darf man den Aufwand für das Management nicht vernachlässigen, welcher auf Path- und Reservierungszuständen anfällt. IntServ arbeitet mit Reservierungen, die über RSVP getätigt werden. IntServ und RSVP sind jedoch getrennte Konzepte. IntServ bietet auch andere Reservierungsmechanismen, welche verwendet werden könnten, andererseits bietet RSVP mehr Signalisierungsmöglichkeiten als für IntServ benötigt werden.

3.2 Differentiated Services (DiffServ)

Beim IntServ-Verfahren werden Ressourcen für einzelne Datenströme belegt, was eine feine Granularität verbunden mit einem sehr hohen Aufwand bedeutet.

Bei DiffServ wird der Verkehr in eine kleine Anzahl von Klassen eingeteilt, denen Ressourcen in heterogenem Ausmaß zugeteilt werden. Man definiert für jede dieser Klassen Handlungsregeln (Per-Hop Behavior, PHB). Die ankommenden Datenpakete werden im Differentiated-Services-Feld des Paket-Headers gemäß ihrer Dienstklasse gekennzeichnet und entsprechend seiner PHB's weiterverarbeitet. Deshalb kommen mehrere unterschiedliche Verkehrsströme mit ähnlichen QoS-Anforderungen zu einem viel größeren Verkehrsbündel zusammen, welches im Netz auf gleiche Weise behandelt wird. Aus diesem Grund ist DiffServ beachtlich einfacher als IntServ. Da nicht jeder Datenfluss einzeln verwaltet werden muss, spart man sich die vielen Zustände und deren Verwaltung im Netz.

Die Vorverarbeitung des Verkehrs, wie z.B. die Markierung der entsprechenden QoS-Klasse oder der Policy-Control, wird nur am Eingang in das DiffServ-Netz vorgenommen. Dazu wird das Differentiated Services Code Point (DSCP)-Feld im DiffServ-Header entsprechend markiert. Für sie kann das Type-of-Service-Feld (TOS)/ Traffic-Class-Octett (TCO)-Feld im IPv4/IPv6-Header genutzt werden. Sie dient dazu, die einzelnen Datenströme eines Verkehrsbündels auseinander zu halten. Das DSCP-Feld hat allerdings 6 Bit Länge und kann daher 64 Klassen darstellen.

Der im DSCP-Feld eingetragene Wert legt dann ein PHB fest. Dies können sein:

- Standard: welches dem Best-Effort-Prinzip entspricht.
- Expedited Forwarding (EF): Pakete mit dieser Markierung sollen möglichst ohne Verzögerung, Latenz und Jitter weitergeleitet werden. Da Paketverluste durch Warteschlangenüberläufe entstehen, sinkt somit die Verlustwahrscheinlichkeit dieser Pakete.

- Assured Forwarding (AF): wurde entwickelt, um einen Standard zur Implementierung verschiedener Stufen von Übermittlungssicherheit von IP-Paketen zur Verfügung zu stellen. Es werden vier Verkehrsklassen definiert, denen unabhängig voneinander Ressourcen (Datenrate, Pufferspeicher) vermacht werden. Innerhalb dieser Klassen darf die Obergrenze der reservierten Datenrate überschritten werden, jedoch in Zeiten von Überlast wird überschüssiger Verkehr verworfen. Dafür wurden in jeder Klasse drei Stufen, welche die relative Wichtigkeit der Pakete innerhalb einer Klasse kennzeichnen, eingeführt, die für das Verwerfen dieses nicht regelkonformen Verkehrs zuständig sind.

4 QoS in Funknetzen

Die Datenübertragung per Funk unterscheidet sich gravierend von der in leitungs-basierten Netzen.

Funkkanäle sind, beschränkt durch die physikalischen Eigenschaften des Übertragungsmediums (Luft), unter anderem wegen den zahlreichen Verbindungsabbrüchen, nicht als zuverlässige Datenübertragungswege zu betrachten. Wegen einiger Eigenschaften ihrer Ausbreitung, wie die Reflexion und Beugung, ermöglichen Funkwellen den Empfang auch außerhalb eines direkten Sichtbereiches. Jedoch genau diese Eigenschaften verursachen Störimpulse, welche bei digitaler Information durch hohe Bit-Fehlerraten bei der Übertragung gezeichnet sind.

Noch einsichtiger zeigt sich dies am Beispiel eines Zieles, welches sich in Bewegung befindet. Funkschatten und Feldstärkenschwankungen führen zum vollständigen Verbindungsabbruch bzw. Signalauslöschung. Die Übertragung wird durch die Störfaktoren erheblich beeinflusst, deswegen sind aus der Sicht der Datenübermittlung Funkverbindungen als unzuverlässige Verbindungen einzustufen.

Alle für den LAN-Einsatz kreierte Anwendungen gehen davon aus, dass eventuell kurz auftretende Unterbrechungen im Sekundenbereich wieder korrigiert werden können. Anderes wie in den Leitungsnetzen, wo Verbindungsabbrüche äußerst selten sind, kann ein Abbruch der Datenübertragungs-Sitzung für Mobilfunkanwendungen wegen ihrer Häufigkeit nicht akzeptiert werden. Die Sitzung muss in diesem Fall aufrechterhalten werden, um ein korrektes Verhalten der beteiligten Programmsysteme zu ermöglichen. Es ist also nicht von einem Abbruch, sondern lediglich von einem vorübergehenden Defekt die Rede.

4.1 Probleme bei der Integration von DiffServ und IntServ

Eine Dienstgütearchitektur, welche DiffServ und IntServ in drahtlosen Netzen repräsentiert, muss folgende Eigenschaften einbeziehen:

1. Hohe Paketverlustraten
2. Niedrige Bandbreite
3. Akkubeschränkung
4. Mobilität

4.1.1 Angepasste Weiterentwicklung von IntServ und DiffServ für mobile Systeme Neben den typischen Dienstgüteparametern wie Jitter, Paketverzögerung, Verlustrate und Bandbreite gibt es bei drahtlosen Verbindungen zusätzliche Parameter, welche die Güte eines Dienstes beschreiben.

1. Verluste in drahtlosen Netzen: Datenverlust ist in mobilen Systemen ein wichtiger Begriff. Dieser kann auftreten während der mobile Endbenutzer die Funkzelle wechselt. Die Gründe dafür können die begrenzte Bandbreite des Übertragungsmediums, das Verwerfen der Pakete und der Ausfall sein. Es wäre von Vorteil für die Anwendung, wenn beschrieben werden könnte, in welcher Art und Weise mit diesem Verlust umzugehen ist. Es werden zwei unterschiedliche Vorschläge diskutiert:
 - Gehäufter Verlust, z.B. bei Tonübertragung
 - Verteilter Verlust, z.B. bei Videoübertragung
2. Akkubeschränkung: Mobile Geräte besitzen Akkus und Batterien, d.h. dass ihre Energie begrenzt ist, deswegen muss sorgfältig darauf geachtet werden. Das Versenden und Empfangen von Paketen kostet Energie, deswegen muss die Basisstation unterrichtet werden, falls der Status der batterie zu niedrig sein sollte.
3. Mobilität: wechselt ein Benutzer die Funkzelle, so müssen evtl. bereits im Voraus in der benachbarten Funkzelle Kapazitäten reserviert werden, um eine verminderte Dienstgüte während oder nach dem Handover zu vermeiden.

4.1.2 IntServ - Modifikationen Speziell beim IntServ-Modell müssen folgende Eigenschaften gesondert betrachtet werden. Durch die Verwendung von RSVP werden viele Auflagen auf das drahtlose mobile Netz gestellt. Ressourcen, welche auf dem ganzen Weg zwischen Sender und Empfänger reserviert werden, können dementsprechend bei einer Änderung der Route auf alte Reservierungen nicht zugreifen sondern benötigen einen Reservierungsmechanismus für Ressourcen.

1. Bandbreitebeschränkung: RSVP sendet periodisch Kontrolldaten um Änderungen im System zu prüfen. Durch die Verwendung eines schmalbandiges Mediums kann dies leicht zu Verstopfung und Überlastung führen. Ist jedoch nur ein Hop zwischen der Basisstation und dem mobilen Endnutzer, ist es nicht mehr notwendig, diese so oft zu verschicken.
2. Energiebeschränkung: Minimale Benutzung der Luftschnittstelle erfordert auch weniger Energie.
3. Mobilität: Mobilität bezieht die Notwendigkeit ein, in Vorfeld Reservationen bei benachbarter Basisstationen zu machen. Hierdurch entstehen folgende Nachteile:
 - (a) Ressourcen sind nicht erhältlich
 - (b) Die Daten können im Verzug oder verloren sein während des Aufbaus der neuen Reservierung.

4.1.3 DiffServ - Modifikationen Speziell beim DiffServ-Modell müssen folgende Eigenschaften gesondert betrachtet werden:

1. Notwendigkeit eines simplen Benachrichtigungsprotokolls: Zur dynamischen Belegung von Ressourcen und Statusmeldungen (z.B. über den aktuellen Akku-Ladestand) benötigt das DiffServ-Modell ein zusätzliches Benachrichtigungsprotokoll. Denkbar ist hier z.B. eine vereinfachte Version des ICMP (Internet Control Message Protokoll).
2. Klassifikation von Datenpaketen innerhalb eines Stroms: wird möglich z.B. durch die Angabe einer zusätzlichen Priorität innerhalb des DSCP (Differentiated Services Code Point) Feldes im Kopf des Pakets.

4.2 Datenübertragungsprotokolle für Funkanwendungen

4.2.1 TCP/IP in mobilen Systemen Das TCP/IP-Protokoll muss in seinen Parametern für die Luftschnittstelle angepasst werden. Die Timeoutbedingungen und Sendefenster können sonst leicht dazu führen, dass nach einem Handover oder Fehler-Burst eine Sendestation nur noch mit sehr kleinen Sendefenstern arbeitet und die eigentlich verfügbare Bandbreite nicht vollständig ausgenutzt wird.

Eine Trennung der logischen Sitzung, das heißt der eindeutig identifizierte und ununterbrochene Datenstrom eine EDV-Anwendung, von der physikalischen, also dem möglicherweise temporär gestörten Übertragungsmedium und seinen Backbone-Protokollen, ist nur möglich, wenn dieses dynamische Verbindungsmanagement oberhalb der Netzwerkinfrastruktur integriert wird. Dann können die vom Netzwerk bereitgestellten Qualitätsparameter zyklisch ausgewertet und in Optimierungsregeln abgebildet werden.

1. Ein mobiles TCP-Protokoll muss berücksichtigen [3]:
 - regelmäßige Verbindungsunterbrechungen
 - kleine Bitraten
 - hohe Bitfehlerraten
2. Das Protokoll muss nach einer Verbindungsunterbrechung schnell wieder mit voller Geschwindigkeit senden können.
3. Die TCP Ende-zu-Ende Semantik soll erhalten bleiben.

4.2.2 Global System for Mobile Communication (GSM) In GSM-Netzen sind die Daten-Übertragungsraten bislang auf 9,6 kBit/s beziehungsweise 14,4 kBit/s im Compressed Mode begrenzt. Moderne multimediale Dienste sind langsam und mit hohen Kosten verbunden.

Durch die in GSM-Netzen üblichen zeitgesteuerten Mehrfachzugriffsverfahren (TDMA - Time Division Multiple Access) wird die Bandbreite eines Übertragungsweges aufgeteilt, indem sie in Zeitschlitze unterteilt wird. In jedem dieser Zeitschlitze wird die volle Bandbreite der Funkstrecke nur einem Nutzer bereitgestellt. Nach Ende seines Zeitschlitzes werden Daten eines anderen Senders übertragen.

Die QoS-Unterstützung bei GSM besteht vor allem aus der Auswahl eines bestimmten Trägerdienstes, der dann eine bestimmte Übertragungsbandbreite zur Verfügung stellen kann.

GSM-Erweiterungen GPRS (General Packet Radio Service) und HSCSD (High Speed Circuit Switched Data) sollen hier weitere Möglichkeiten öffnen.

4.2.3 High Speed Circuit Switched Data (HSCSD) HSCSD entspricht einer GSM-Kanalbündelung. Der Netzbetreiber bestimmt, wie viele Kanäle gleichzeitig dem Nutzer zur Bündelung zur Verfügung stehen. Es sind acht Kanäle vorstellbar, realisiert ist derzeit aber nur die Bündelung von vieren. Alternativ von zwei zu zwei oder drei zu eins stehen für den Down- und Upload zur Wahl. Der Vorteil von HSCSD ist die feste Übertragungsbandbreite, die während der ganzen Verbindung unverändert dem Nutzer zur Verfügung steht. Bei einer hohen Auslastung des Netzes ist es für den Netzbetreiber möglich, Einschränkungen vorzunehmen, indem er Prioritäten setzt. Er kann Sprachtelefonie vorziehen und anderen Diensten Kanäle "klauen".

QoS Unterstützung wird aus der Anzahl der Kanäle abgeleitet.

Es können auch asymmetrische Verbindungen hergestellt werden, z.B. kann für das Surfen im Internet mehr Bandbreite für den Downstream zur Verfügung gestellt werden als für den Upstream.

Folgende Probleme können auftreten:

- Die dauerhafte Belegung der Kanäle, was zu unbenutzter Kapazität führt, welche dann auch keinem anderen Nutzer zur Verfügung stehen
- Die erhöhte Blockierwahrscheinlichkeit, durch die parallel belegte Kanäle.

4.2.4 General Packet Radio Service (GPRS) GPRS ist ein spezielles System für die Übertragung von Informationen über das GSM- Netzwerk mit dem Ziel, eine höhere Übertragungsgeschwindigkeit zu erreichen. Hierbei werden die Daten in einzelne Pakete aufgeteilt, wobei jedes dieser Pakete mit zusätzlichen Informationen gekoppelt wird. Pakete können durch verschiedene Zeitschlitze des Netzwerks geführt werden. Diese stehen mehreren Nutzern gleichzeitig zur Verfügung, was die Nutzung freier Kapazitäten gestattet. Beim Empfänger werden die Pakete in der richtigen Reihenfolge zusammengefügt.

Welche Vorteile bietet uns die paketorientierte Vermittlung? GPRS ermöglicht erstmals die Abrechnung nach übermittelltem Datenvolumen. Anwendungen wie WAP sind damit deutlich günstiger, da hier nur kleine Datenmengen übertragen werden. Internetradio oder Videostreams bereiten dagegen mehr Probleme, da GPRS keine feste Bandbreite garantiert. GPRS eröffnet mit der Always-On-Funktionalität neue Perspektiven. Diese Funktion bietet dem Kunden eine permanente Online-Verbindung, deswegen entfällt die lästige und umständliche Einwahlprozedur in das mobile Funknetz.

In GPRS sind fünf verschiedenen Dienstgüte-Klassen definiert, welche die Übertragungseigenschaften definieren [4].

1. Verzögerungsklasse (Delay Class)
2. Dringlichkeitsklasse (Precedence Class)
3. Verlässlichkeitsklasse (Reliability Class)
4. Spitzendurchsatzklasse (Peak Throughput Class)
5. Durchschnittsdurchsatzklasse (Mean Throughput Class)

4.2.5 Enhanced Data Rates for GSM (EDGE) Es gibt auch andere Möglichkeiten, die Bandbreite zu erhöhen wie z.B. Enhanced Data Rates for GSM (EDGE). Durch diesen Dienst wird erlaubt, dass mehrere Gespräche in einem Zeitschlitz übertragen werden. So wird die Kapazität erhöht, denkbar ist dass Dienste mit höherer Qualität angeboten werden.

Er kann sowohl mit GPRS wie auch HSCSD benutzt werden und stellt eine echte Alternative als Zwischenschritt zu Diensten der 3.Generation dar.

QoS wird durch den entsprechenden Datendienst, und nicht durch EDGE selbst festgelegt.

EDGE könnte auch als Lösung für Mobilfunkbetreiber ohne UMTS Lizenz betrachtet werden, da es sehr hohe Übertragungsrate ermöglicht. Unter Verwendung aller 8 Zeitschlitze und bei guter Verbindungsqualität sind Datenraten bis zu 384 kbit/s möglich.

5 IEEE 802.11e

Heutige Wireless Local Area Networks (WLANs) basieren auf dem Standard 802.11 des Institute of Electrical and Electronics Engineers (IEEE). Die eingesetzten Codierungs- und Zugriffsverfahren ermöglichen Datenübertragungsraten von mittlerweile bis zu 54 Mbit/s (dabei geht jedoch ein großer Teil der Bandbreite durch Kontrolldaten verloren).

Echte Dienstgarantien werden jedoch bisher nicht unterstützt. Zwar ist in 802.11 im Infrastrukturmodus eine gewisse Art der zentralen Koordination gegeben (s. unten), doch dieser stellt noch keine ausreichenden Funktionalitäten für echte Dienstgarantien zur Verfügung. Gerade weil sich WLAN immer mehr verbreitet und die bisherigen 802.11-Standards nur nach dem "Best Effort Prinzip" arbeiten, werden QoS-Standards im Zusammenhang mit den wachsenden Anwendungsanforderungen benötigt werden.

Aktuell beschäftigt sich die Arbeitsgruppe IEEE 802.11e mit diesen Problemstellungen. Doch bevor auf die Erweiterungen und Neuerungen im einzelnen eingegangen wird, folgt eine kurze Übersicht über die bisherigen Kontrollmechanismen zum Zugriff auf die Luftschnittstelle (Medium Access Control - MAC) in 802.11.

5.1 Infrastruktur- und Ad-Hoc-Modus in 802.11

Innerhalb einer Funkzelle (Basic Service Set - BSS) lässt sich der Zugriff auf die Luftschnittstelle bei IEEE 802.11 in zwei Modi unterteilen:

- den Infrastrukturmodus, in dem ein sog. Point Coordinator (PC) den Datenverkehr zwischen allen Stationen (STAs) kontrolliert
- den Ad-Hoc-Modus, in dem es keine zentrale Kontrollinstanz gibt und alle STAs gleichberechtigt miteinander kommunizieren.

Die beiden wesentlichen Funktionen, welche hierbei zum Einsatz kommen sind

- die "Point Coordination Function" (PCF) für den Infrastrukturmodus (welche sich in der Praxis jedoch nicht durchgesetzt hat)
- die "Distributed Coordination Function" (DCF) für den Ad-Hoc-Modus (Abb. 4)

Auf beide wird im Folgenden nicht genauer eingegangen, sie bilden jedoch die Grundlage für die weiteren Ausführungen (eine ausführliche Beschreibung findet sich unter [8]).

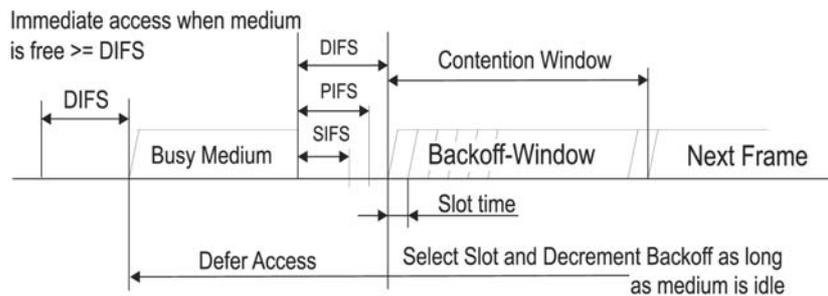


Fig. 4. Distributed Coordination Function (DCF) in 802.11. Quelle: [8]

5.2 Neuerungen in 802.11e - MAC Erweiterungen

Der Schwachpunkt der bisherigen PCF in Bezug auf QoS liegt darin, dass eine Station, die während der konkurrenzfreien Sendephase (Contention Free Period, CFP) vom PC die Erlaubnis zum Senden bekommen hat, ein Datenpaket mit einer Größe von bis zu 2304 Bytes senden darf. Damit hat der PC keine Kontrolle mehr über die genaue Größe der übertragenen Daten und eine exakt gesteuerte Belegung der Luftschnittstelle ist nicht möglich.

Zur Lösung dieses Problems schlägt 802.11e die Einführung zweier neuer Funktionen vor:

- die Enhanced Distributed Coordination Function (EDCF)
- die Hybrid Coordination Function (HCF).

Analog zum PC in bisherigen 802.11-Netzen übernimmt bei 802.11e der sog. Hybrid Coordinator (HC) die Rolle einer zentralen Kontrollinstanz. Der HC kann jedoch zu einer beliebigen Zeit (nach Verstreichen eines PCF Interframe Space - PIFS, s.unten) die Funkschnittstelle belegen und dann dediziert Kapazitäten an die um QoS erweiterten STAs (QSTAs) vergeben. In bisherigen 802.11-Netzen ist die Contention Free Period (CFP) durch den sog. Beacon Frame (BC) festgelegt und der PC kann ein Polling der Client-Rechner nur innerhalb dieser CFP vornehmen (Abb. 5). Die EDCF ist Teil der HCF und ermöglicht weiter-

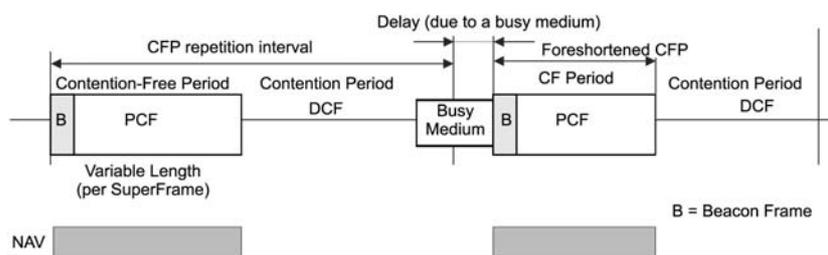


Fig. 5. DCF und PCF im Wechsel. Quelle: [8]

hin einen Ad-Hoc-Betrieb, wurde jedoch im Vergleich zur bisherigen DCF insofern erweitert, als das ein gewisses Maß an Dienstgüte durch die Einführung von Verkehrsklassen (Traffic Categories - TCs) zur Verfügung gestellt wird.

Eine weitere wichtige Neuerung in 802.11e ist, dass die MAC-Empfangsbestätigung (ACK) optional ist. Besonders bei Anwendungen, deren Daten nur eine sehr kurze Time-To-Live (TTL) besitzen, z.B. bei Voice over IP (VoIP), werden dadurch erheblich bessere Ergebnisse erzielt.

5.2.1 Enhanced Distributed Coordination Function (EDCF) EDCF erweitert die bestehende DCF und bietet QoS nach dem dringlichkeitsbasierten Best Effort Prinzip, vergleichbar mit dem DiffServ-Modell.

Eine QSTA versucht dabei wie bei der bisherigen DCF, Zugriff auf die Funkschnittstelle zu bekommen - sie konkuriert mit den anderen QSTAs um eine sog. Transmission Opportunity (TXOP). Eine TXOP ist ein begrenztes Zeitintervall, in welchem eine QSTA den Zugriff auf die Funkschnittstelle zur Übermittlung einer Folge von Frames zugesichert bekommt. Eine TXOP wird durch eine Startzeit und eine Maximaldauer definiert.

Zusätzlich verwaltet jede QSTA intern bis zu acht virtuelle Stationen mit jeweils individuellen Back-Off-Zählern entsprechend der TCs. Erreichen die Zähler zweier paralleler TCs innerhalb einer QSTA zur gleichen Zeit Null, so kommt

es zu einer virtuellen Kollision und der interne Scheduler gibt die TXOP an die TC mit der höchsten Priorität (siehe Fig. 6). Die unterschiedliche Priorisierung

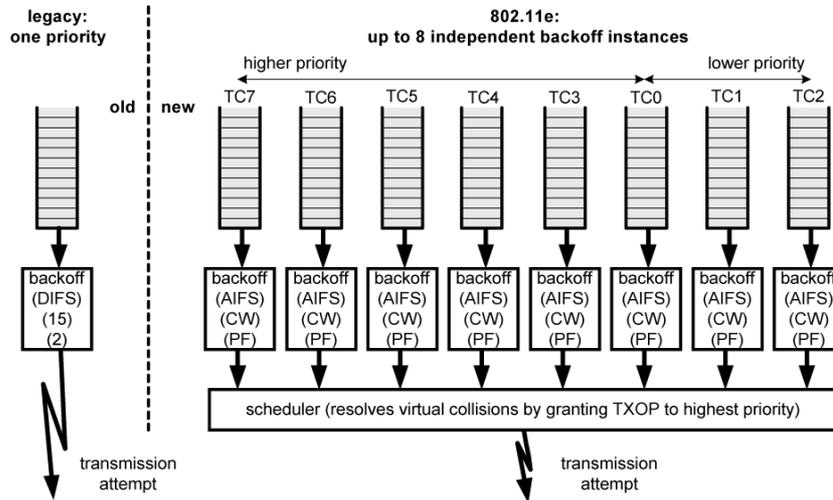


Fig. 6. Virtuelle Back-Off-Zähler. Quelle: [10]

der TCs wird durch die Einführung sog. Arbitrary Inter Frame Spaces (AIFSs) realisiert. Dabei gilt: je höher die Priorität einer TC[i] ist, desto kürzer ist ihr zugehöriger AIFS[i]. Zusätzlich gilt immer:

$$SIFS \leq PIFS \leq DIFS \leq AIFS[i]$$

Fig. 8 zeigt die AIFSs dreier TCs. Jede TC[i] ($i \in \{1, 2, \dots, 8\}$) beginnt, ihren individuellen Back-Off-Zähler zu dekrementieren, nachdem die Funkschnittstelle für mindestens den entsprechenden AIFS[i] frei gewesen ist. Das Vergrößern des Back-Off-Zählers nach dem Auftreten einer Kollision geschieht nicht wie bei der herkömmlichen DCF (s. Fig. 7) durch Verdoppeln des Konkurrenzintervalls (Contention Window, CW), sondern ergibt sich aus:

$$neuesCW[i] = ((altesCW[i] + 1) * PF) - 1,$$

wobei PF ein Beständigkeitsfaktor (Persistence Factor, PF) ist, der bestimmt, wie schnell CW[i] wächst. $PF = 2$ entspricht dem herkömmlichen Verfahren in 802.11. Hat eine TC vom internen Scheduler die TXOP erhalten, so versucht diese, nach dem Prinzip der DCF ihre Daten zu senden. Hierbei kann es wie bisher ebenfalls zu Kollisionen kommen, entsprechend wird dann der Back-Off-Zähler dieser TC erhöht.

Eine höher priorisierte TC kann so (im Vergleich zu einer TC mit niedrigerer Priorität) durch ihren kürzeren AIFS öfters ihren Back-Off-Zähler dekrementieren und öfters Daten übertragen.

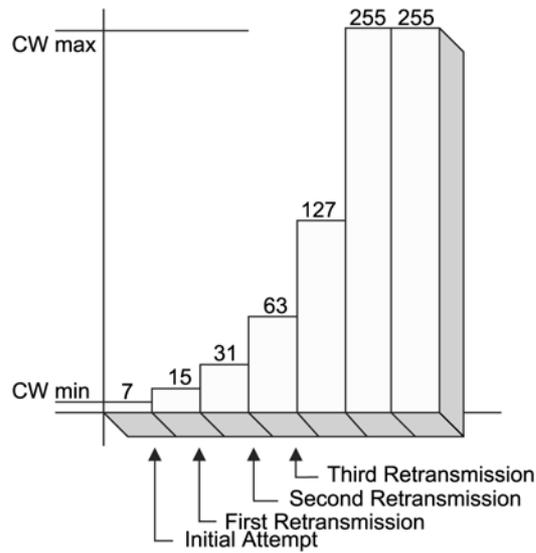


Fig. 7. Vergrößerung des Contention Fensters mit Beständigkeitsfaktor $PF = 2$. Quelle: [8]

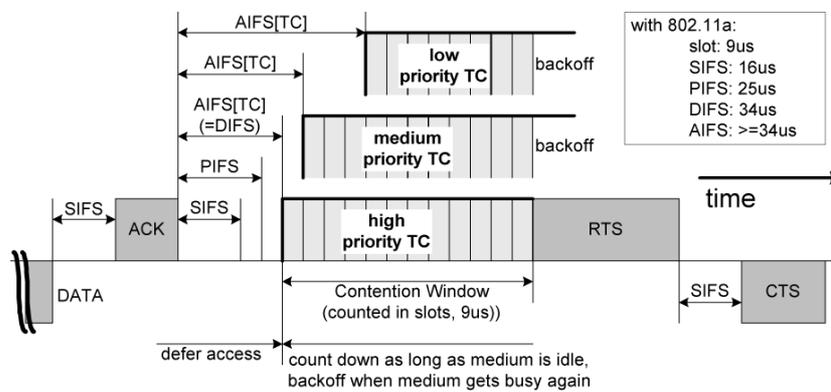


Fig. 8. Traffic Categories (TCs) und AIFSS. Quelle: [10]

Innerhalb der Contention Period (CP) belegte TXOPs (EDCF-TXOPs) sind in ihrer Größe durch die Festlegungen im periodisch verbreiteten Beacon Frame (BF) begrenzt. TXOPs, welche durch die HCF zugeteilt werden (polled-TXOPs), spezifiziert der HC. [10] zeigt, dass mit ECDF eine gewisse Dienstgüte erreicht werden kann. Die Wahrscheinlichkeit, dass eine TC mit niedriger Priorität senden darf ist statistisch gesehen kleiner als bei einer TC mit hoher Priorität.

Die Ergebnisse hängen jedoch stark von den gewählten Werten für PF, CWMin[i] und CWMax[i] (minimale und maximale Länge der CWs), EDCF-TXOP und AIFS[i] ab.

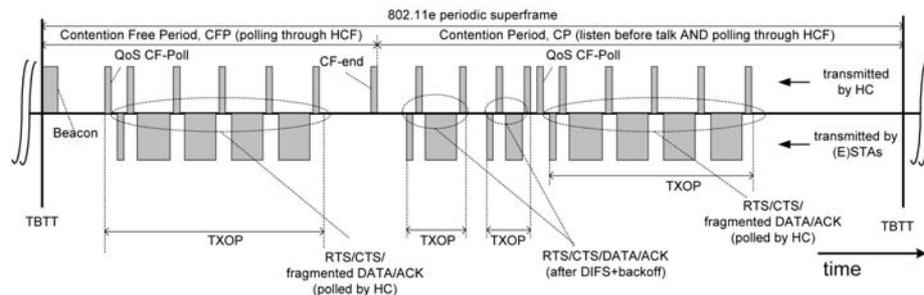


Fig. 9. Ein 802.11e Superframe. Quelle: [10]

5.2.2 Hybrid Coordination Function (HCF) HCF erweitert die EDCF. Der HC kann jederzeit entweder sich selbst oder anderen QSTAs dediziert TXOPs erteilen, nachdem die Luftschnittstelle mindestens für einen PIFS frei gewesen ist.

CP und CFP bilden zusammen einen sog. Superframe (s. Fig 9). Während der CFP versucht keine Station, selbstständig Zugriff auf die Funkschnittstelle zu bekommen, sondern jeder Station wird die Erlaubnis zum Senden vom HC durch einen QoS CF-Poll Frame erteilt. Die CFP endet mit einem CF-End Frame oder nach der im BF angegebenen Zeit. Hier zeigt sich nun der Vorteil gegenüber der herkömmlichen PCF: der HC hat Kontrolle über die Länge des Zeitintervalls (TXOP), dass eine Station zum Senden zugewiesen bekommt.

Die QSTAs teilen dem HC ihre Dienstgütereorderungen in speziellen QoS-Management-Frames mit. Dabei wird für jeden benötigten Datenstrom (traffic stream, TS) eine Verkehrsspezifikation (traffic specification, TSPEC) angegeben. Der HC berechnet aus allen TSs einer Station ihre Anforderungen und teilt entsprechend die TXOPs zu. Die wesentlichen Parameter einer TSPEC sind in Tabelle 1 angegeben mit Beispielwerten für VoIP und Video (MPEG-4) [7].

TSPEC-Parameter	VoIP (G.729A)	Video (MPEG-4)
Mean Data Rate	24 kb/s	630 kb/s
Delay Bound	60 ms	60 ms
Nominal MSDU Size	60 octets	1024 octets
Maximum MSDU Size	60 octets	1024 octets
Maximum burst size	120 octets	14894 octets
Peak data rate (in bits per second)	24 kb/s	1,5 Mb/s
Minimum PHY rate (in bits per second)	24 Mb/s	24 Mb/s

Table 1. TSPEC-Parameter in 802.11e [7]

5.3 Bemerkungen und offene Fragen

Besonders bei der Integration von neuen 802.11e-STAs in bestehende WLANs stellt sich die Frage nach Kompatibilität und Effizienz. Operieren STAs und QSTAs innerhalb einer Zelle, so erlangen die QSTAs höhere Übertragungsraten, falls für CWMin kleine Werte (≈ 5 Zeitschlitze) gewählt werden [10]. Noch nicht geklärt ist das Problem von sich überlappenden Zellen, die unter 802.11e operieren (Overlapping QBSS). Hier sind weitere Untersuchungen nötig.

6 QoS in UMTS

Bereits 1992 wurde von der ITU (International Telecommunications Union) der IMT-2000-Standard (International Mobile Telecommunications at 2000MHz) ins Leben gerufen, um international eine Vereinheitlichung für Mobilfunksysteme im Frequenzbereich um 2000 MHz zu schaffen. Das Universal Mobile Telecommunication System (UMTS) ist der Europäische und ITU-2000-konforme Standard für Mobilfunknetzte der 3. Generation.

UMTS ermöglicht insbesondere die Integration und Weiterverwendung bestehender GSM-Netze. Im Gegensatz dazu kommt z.B. in Amerika der CDMA2000-Standard zum Einsatz, der die Integration des dortigen IS-95 Mobilfunknetzes ermöglicht.

Die Ziele von UMTS sind zum einen die effizientere Ausnutzung der Frequenzen und dadurch schnellere Datenraten (es sollen Übertragungsraten von bis zu 2 Mbps für den Endnutzer möglich werden), zum anderen soll den Benutzern differenzierte Dienstgüte (je nach Anwendungsanforderungen) garantiert werden können.

6.1 UMTS-QoS-Klassen

Das 3rd Generation Partnership Project (3GPP) definiert in seiner technischen Spezifikation TS 23.107 (Release 5) die "QoS Konzepte und Architekturen" für UMTS [6]. Die Spezifikation unterscheidet vier verschiedene QoS-Klassen:

- Konversationsklasse (conversational class)
- Streaming Klasse(streaming class)
- Interaktive Klasse (interaktive class)
- Hintergrundübertragung (background class)

Die Klassen unterscheiden sich besonders hinsichtlich ihrer Anforderungen gegenüber der Verzögerung (Delay).

Konversationsklasse: Echtzeitkonversation (z.B. Telefonie, Videokonferenz)

Eigenschaften: geringe Verzögerung, geringer Jitter (Fluktuation im Datenstrom)

Die Anforderungen der Konversationsklasse richten sich im wesentlichen nach der menschlichen Wahrnehmungsfähigkeit [5]. Z.B. gibt die ITU-T Recommendation G.114 für die Einweg-Verzögerung bei Sprachübertragung einen Wert zwischen 0 und 150 ms als bevorzugten, zwischen 150 und 400 ms als akzeptablen und über 400 ms als inakzeptablen Bereich an. Verzögerungen unterhalb von 30 ms werden vom menschlichen Gehör nicht wahrgenommen. Noch empfindlicher reagiert das Gehör auf Jitter: ITU-T schlägt einen Wert um 1 ms vor. Verzerrungen im Signal sind ebenfalls akzeptabel, solange die die Rahmenverlustrate (Frame Erasure Rate, FER) 3% nicht übersteigt.

Streaming Klasse: Video- oder Audio-Datenströme (unidirektional)

Eigenschaften: konstante Übertragungsgeschwindigkeit

In die Streaming Klasse fallen unidirektionale Datenströme zur Video- oder Audioübertragung, deren Empfänger eine natürliche Person ist. Die akzeptablen Wertebereiche für Jitter und Verzögerung sind wesentlich größer als in der Konversationsklasse, da verzerrte oder verzögerte Bereiche in einem Datenstrom i.d.R. von der empfangenden Applikation durch eine Funktion wieder korrekt ausgerichtet werden. Die Leistungsfähigkeit dieser Funktion bestimmt die akzeptablen Werte für Jitter und Verzögerung.

Interaktive Klasse: Mensch-Maschine- oder Maschine-Maschine-Interaktion

Eigenschaften: keine fehlerhaften Übertragungen

Anwendungen wie z.B. der Zugriff auf Datenbanken und www-Seiten oder remote Login fallen in die Streaming Klasse. Aber auch das Versenden einer eMail geschieht innerhalb der Interaktiven Klasse, nämlich dann, wenn der Mail User Agent (MUA) des Benutzers die eMail an den Mail Transfer Agent (MTA) der Benutzers (i.d.R. ein Internet Service Provider) zustellt. Akzeptable Zeiten liegen hierbei zwischen 2 bis 4 Sekunden.

Hintergrundübertragung: Hintergrunddatenverkehr

Eigenschaften: keine fehlerhaften Übertragungen, der Empfänger erwartet die Daten nicht innerhalb einer bestimmten Zeit.

Bei der Zustellung von eMail oder SMS oder auch Faxübertragung geschieht dies innerhalb der Klasse für Hintergrundübertragung. Für SMS oder Fax ist eine Verzögerung von bis zu 30 Sekunden akzeptabel, bei der eMailzustellung (von Server zu Server) gehen die Vorstellungen der Benutzer weit auseinander, der Durchschnittswert liegt bei mehreren Stunden.

Bemerkung: In seinen Spezifikationen behandelt 3GPP ab Release 5 nur noch rein paketvermittelte Netze. Alle leitungsvermittelten Dienste sind hinsichtlich der obigen Klasseneinteilung immer der Konversationsklasse zugeordnet.

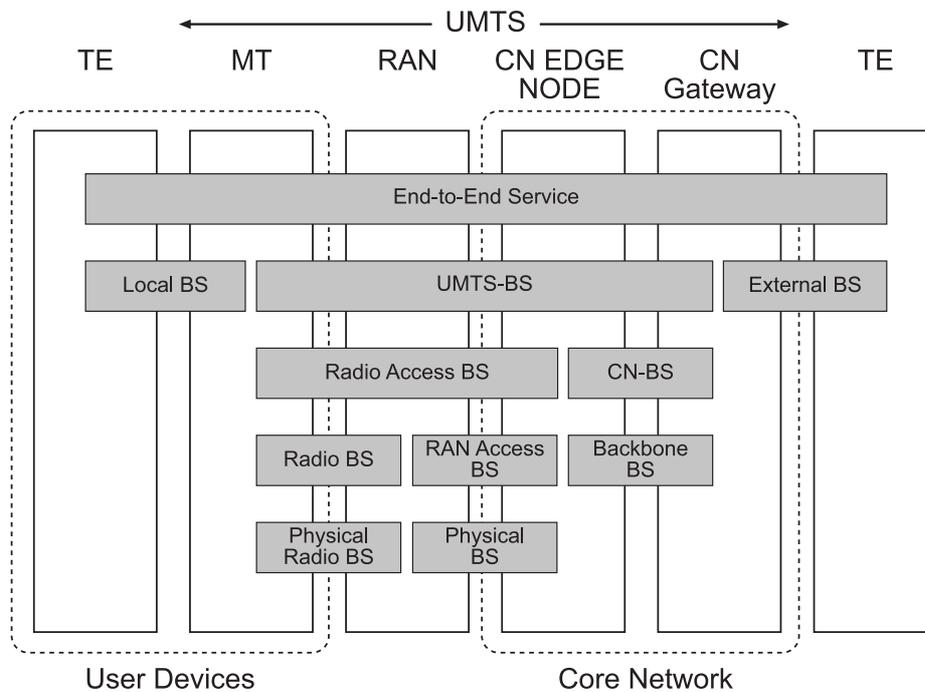


Fig. 10. Die UMTS-QoS-Architektur. TE = Terminal Equipment, MT = Mobile Termination, RAN = Radio Access Network, CN = Core Network, BS = Bearer Service

6.2 Die UMTS QoS - Architektur

Die UMTS QoS-Architektur ist in mehrere Schichten unterteilt (s. Fig. 10). An oberster Stelle steht die Schicht des **Ende-zu-Ende Dienstes**. Hier wird zwischen zwei Datenendgeräten (Terminal Equipment, TE) eine Dienstgüte ausgehandelt (Konversations-, Streaming-, interaktive Klasse oder Hintergrundübertragung).

Eine Schicht darunter werden drei verschiedene Trägerdienste (Bearer Service, BS) unterschieden:

1. der **lokale BS** zwischen dem Datenendgerät eines Benutzers und einem Mobilanschluss an das UMTS-Netz (z.B. ein Notebook, das via Infrarotverbindung mit einem UMTS-Handy verbunden ist),
2. der **externe BS** zwischen dem Gateway des Kernnetzes (Core Network Gateway, CNGW) und einem Datenendgerät außerhalb des UMTS-Netzes (z.B. die Verbindung zwischen dem CNGW und einem VoIP-Telefon oder dem Webserver eines ISPs),
3. der **UMTS-Trägerdienst (UMTS-BS)**.

Der lokale und der externe Trägerdienst liegen außerhalb des Zuständigkeitsbereiches des UMTS-Netzes und werden daher hier nicht weiter behandelt.

QoS wird in UMTS durch den UMTS-BS bereitgestellt. Auf dieser Schicht wird QoS zwischen dem UMTS-Netz und externen Komponenten ausgehandelt. Der UMTS-BS verhandelt die von der Ende-zu-Ende-Schicht angeforderte Dienstgüte auf der darunter liegenden Schicht mit dem **Radio Access- (RABS)** und dem **Core Network-Trägerdienst (CN-BS)**. Der CN-BS ist zuständig für die Überprüfung der Benutzerdaten und die Verbindung zwischen CN Randknoten und dem CNGW über das UMTS Backbone Netz. Der RABS verwaltet den Zugriff auf das Radio Access Network (RAN) und überprüft, ob für die geforderte Dienstgüte entsprechende Ressourcen im RAN zur Verfügung stehen.

Unterhalb des RABS liegen zum einen **Radio Bearer Service (RBS)** und **Physical Radio Bearer Service (PRBS)**, welche für die Kommunikation zwischen einem MT und RAN über die Funkschnittstelle zuständig sind, zum anderen **RAN Access BS** und **Physical Bearer Service (PBS)**, diese sorgen für den Transport zwischen RAN und CN. RBS und PRBS werden z.B. durch UTRAN FDD (UMTS Terrestrial Radio Access - Frequency Division Duplex) oder UTRAN TDD (- Time Division Duplex) bereitgestellt.

6.2.1 Trägermanagement in UMTS Wie wird nun eine bestimmte Dienstgüte zwischen einem Datenendgerät und dem UMTS-Netz ausgehandelt? Fig. 11 zeigt schematisch, welche Nachrichten dabei ausgetauscht werden.

Die wichtigsten QoS-Parameter in UMTS sind:

- maximale Bitrate (kb/s)
- garantierte Bitrate (kb/s)
- erlaubte Verzögerung/Delay (ms)
- ob über die angeforderten QoS-Parameter verhandelt werden darf

Zuerst ordnet die für das Kommunikationsmanagement (communication management, CM) des Benutzergerätes (User Equipment, UE) zuständige Einheit dem benötigten Dienst entsprechende QoS-Parameter zu. Das UE initiiert daraufhin einen Verbindungsaufbau und über das UTRAN wird in einer Radio Ressource Control (RRC) Nachricht eine eingekapselte "UE initial message" verschickt, welche die gewünschten QoS-Parameter enthält. Zwischen UTRAN und CN wird Nachricht in einer Radio Access Network Application Part (RANAP) Nachricht transportiert.

Im CN ordnet die CM-Einheit die Anfrage entsprechend der übergebenen QoS-Parameter einem UMTS-BS zu. Der Radio Resource Management (RRM) Admission Control (AC) -Algorithmus überprüft nun, ob die geforderten Parameter momentan physikalisch erfüllbar sind. Sind keine oder nur unzureichende Kapazitäten vorhanden, so werden die QoS-Parameter neu ausgehandelt (sofern das UE dies zugelassen hat) oder die Anfrage wird abgelehnt oder kommt in eine Warteschlange. Andernfalls kommt eine Verbindung zustande und der Benutzer bekommt eine Verbindung mit der angeforderten bzw. ausgehandelten Dienstgüte zugeteilt.

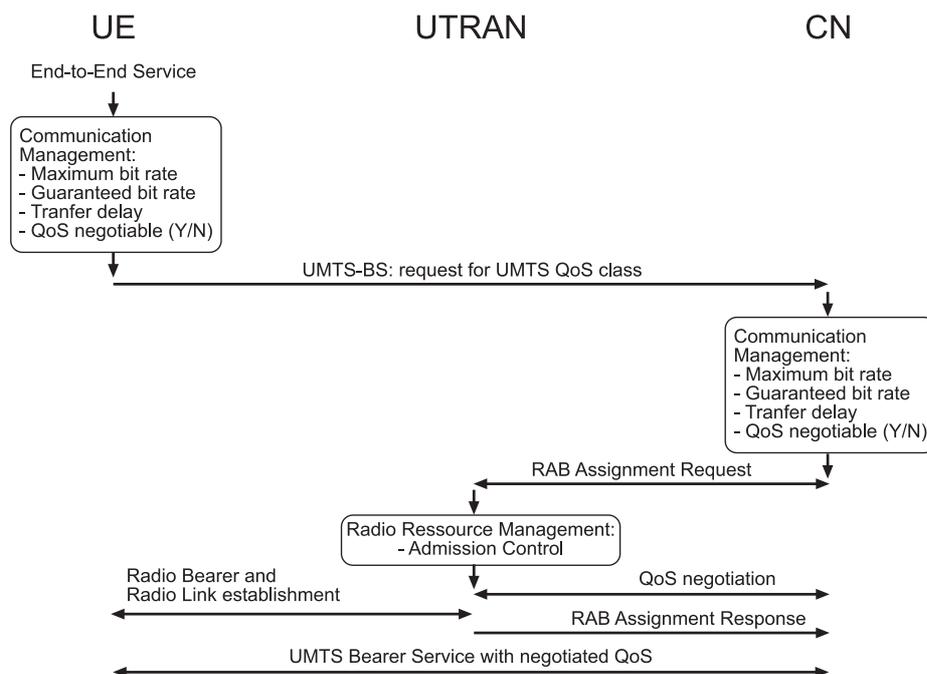


Fig. 11. Trägermanagement in UMTS. Quelle: [9]

6.3 UMTS-QoS-Attribute

Der UMTS-Trägerdienst unterscheidet folgende Attribute:

- Verkehrsklasse (konversation, streaming, interaktiv, hintergrund)
Art der Anwendung. Mit Hilfe dieses Parameters kann UMTS Annahmen über die Datenquelle machen und den Transport entsprechend optimieren.
- maximale Bitrate (kbps)
Maximale Anzahl der an einem Dienstzugangspunkt (Service Access Point, SAP) ausgelieferten Bits innerhalb eines bestimmten Zeitintervalls, geteilt durch die Länge des Zeitintervalls.
- garantierte Bitrate (kbps)
Garantierte Anzahl der an einem Dienstzugangspunkt ausgelieferten Bits innerhalb eines bestimmten Zeitintervalls (vorausgesetzt es liegen Daten zum Senden vor), geteilt durch die Länge des Zeitintervalls.
- Auslieferungsreihenfolge (ja/nein)
Gibt an, ob der UMTS-Trägerdienst die SDUs entsprechend der Quell-Reihenfolge beim Empfänger ausliefern soll.
- maximale SDU (Service Data Unit)-Größe (octets)
- SDU Formatierungsinformation (bits)
Gibt die exakte Größe der SDUs an. Ist eine Anwendung in der Lage, diesen Parameter mit anzugeben, so kann der Trägerdienst effizienter operieren.
- Anteil fehlerhafter SDUs
Nur anwendbar in den Klassen, die fehlerhafte SDUs zulassen.
- Anteil der unentdeckt fehlerhaft übertragenen Bits
Nur anwendbar in den Klassen, die fehlerhafte SDUs zulassen.
- Auslieferung fehlerhafter SDUs (ja/nein/-)
"ja": Fehlererkennung aktiv, fehlerhafte SDUs werden mit einer entsprechenden Kennzeichnung übermittelt.
"nein": Fehlererkennung aktiv, fehlerhafte SDUs werden verworfen.
"-": alle SDUs werden ohne Fehlerüberprüfung übermittelt.
- Transferverzögerung (ms)
Gibt die maximale Verzögerung für 95% aller auf dem Träger übertragenen SDUs an und bezieht sich auf die Zeit, die eine SDU von einem SAP bis zur Abgabe bei einem anderen SAP benötigt.
- Traffic handling priority
Gibt die Wichtigkeit aller diesem Träger zugehörigen SDUs relativ zu SDUs anderer Träger an.
- Bereitstellung-/Zurückhaltungspriorität (allocation/retention priority)
Dieser Parameter ist ein Teilnehmerattribut und kann nicht vom MT ausgehandelt werden. Ist z.B. das Netz zeitweise stark ausgelastet, so kann hiermit bestimmten Nutzern (z.B. bevorzugten Kunden) der Vorrang gewährt werden.
- source statistics descriptor (Sprache / unbekannt)
Sprache folgt in ihrem Datenaufkommen bestimmten statistischen Mustern. Bekommt das UMTS-Träger die Information, dass es sich bei den übertragenen

SDUs um Sprachdaten handelt, so kann die Übertragung mit Hilfe spezieller Algorithmen zusätzlich optimiert werden.

- signalling indication (yes/no)

Ist nur für die interaktive Klasse definiert und gibt an, ob es sich um spezielle Signalisierungsnachrichten handelt, die ggf. bevorzugt behandelt werden soll.

In Fig. 12 sind die von 3GPP vorgeschlagenen Wertebereiche aufgelistet.

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate (kbps)	$\leq 16\ 000$	$\leq 16\ 000$	$\leq 16\ 000$ - overhead	$\leq 16\ 000$ - overhead
Delivery order	Yes/No	Yes/No	Yes/No	Yes/No
Maximum SDU size (octets)	$\leq 1\ 500$ or $1\ 502$	$\leq 1\ 500$ or $1\ 502$	$\leq 1\ 500$ or $1\ 502$	$\leq 1\ 500$ or $1\ 502$
Delivery of erroneous SDUs	Yes/No/-	Yes/No/-	Yes/No/-	Yes/No/-
Residual BER	$5 \cdot 10^{-2}, 10^{-2}, 5 \cdot 10^{-3}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$	$5 \cdot 10^{-2}, 10^{-2}, 5 \cdot 10^{-3}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$	$4 \cdot 10^{-3}, 10^{-5}, 6 \cdot 10^{-6}$	$4 \cdot 10^{-3}, 10^{-5}, 6 \cdot 10^{-6}$
SDU error ratio	$10^{-2}, 7 \cdot 10^{-3}, 10^{-3}, 10^{-4}, 10^{-5}$	$10^{-1}, 10^{-2}, 7 \cdot 10^{-3}, 10^{-3}, 10^{-4}, 10^{-5}$	$10^{-3}, 10^{-4}, 10^{-6}$	$10^{-3}, 10^{-4}, 10^{-6}$
Transfer delay (ms)	100 – maximum value	280 – maximum value		
Guaranteed bit rate (kbps)	$\leq 16\ 000$	$\leq 16\ 000$		
Traffic handling priority			1,2,3	
Allocation/Retention priority	1,2,3	1,2,3	1,2,3	1,2,3
Source statistic descriptor	Speech/unknown	Speech/unknown		
Signalling Indication			Yes/No	

Fig. 12. UMTS-BS-Attribute. Quelle: [6]

7 Zusammenfassung

Vergleicht man den UMTS-Standard mit den vorgestellten Vorschlägen zu IEEE 802.11e, so fällt vor allem auf, dass bei UMTS nur die QoS-Parameter und deren zulässige Wertebereiche vorgegeben werden. Die konkrete Implementierung bleibt dem Hersteller überlassen und wird nicht näher spezifiziert. 802.11e hingegen gibt exakt vor, welche Verfahren zur Realisierung von QoS eingesetzt werden und wie die Zugriffsfunktionen auf der MAC-Schicht konkret aufgebaut sind.

Weiterhin unterscheiden sich UMTS und der neue WLAN-Standard grundlegend in ihrer Art des Zugriffs auf die Luftschnittstelle. Bei WLAN war Dienstgüte beim Entwurf des ursprünglichen 802.11-Standards nicht eingeplant, so dass die ECDF in 802.11e letztendlich auf einem statistischen Zugriffsverfahren aufsetzt. Bei UMTS wurde QoS bereits in der Entwicklungsphase mit eingeplant, die Ressourcen können fest zugeteilt bzw. zwischen den Komponenten ausgehandelt werden.

Der 802.11e-Standard befindet sich aktuell noch in der Entwicklungsphase und ob er sich in der Praxis durchsetzt, wird sich erst zeigen. UMTS steht (zumindest in Europa) erst am Anfang der praktischen Umsetzung, auch hier

ist zum Teil noch offen, in welchem Umfang die vorgeschlagene QoS-Architektur umgesetzt und wie sie von den Benutzern aufgenommen werden wird.

References

- [1] R. Steinmetz, Multimedia-Technologie, Springer, Berlin, (2000)
- [2] Eduard Siemens, Quality of Service in Rechnernetzen, Universität Hannover, 1999
- [3] <http://www.hegering.informatik.tu-muenchen.de/Events/Sarntal/Sarntal2000.html>
- [4] http://umtslink.at/GPRS/qos_gprs.htm
- [5] 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects: "Service aspects: Services and Service Capabilities" (Release 5)
- [6] 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects: Quality of Service (QoS) concept and architecture (Release 5)
- [7] Antonio Grilo, Mario Macedo, Mario Nunes, INESC/IST: A Scheduling Algorithm for QoS Support in 802.11e Networks, IEEE Wireless Communications, June 2003, Vol. 10 No. 3, pp 36 - 43
- [8] ANSI/IEEE Std 802.11, 1999 Edition
- [9] Heikki Kaaranen, Ari Ahtiainen, Lauri Laitinen, Siamak Ghian, Vallteri Niemi: "UMTS Networks: Architecture, Mobility and Service", pp. 144 - 151, Wiley-Verlag, ISBN 047148654
- [10] Stefan Mangold, Sunghyun Choi, Peter May, Ole Klein, Guido Hiertz, Lothar Stibor: IEEE 802.11e Wireless LAN for Quality of Service, <http://citeseer.nj.nec.com/537394.html>