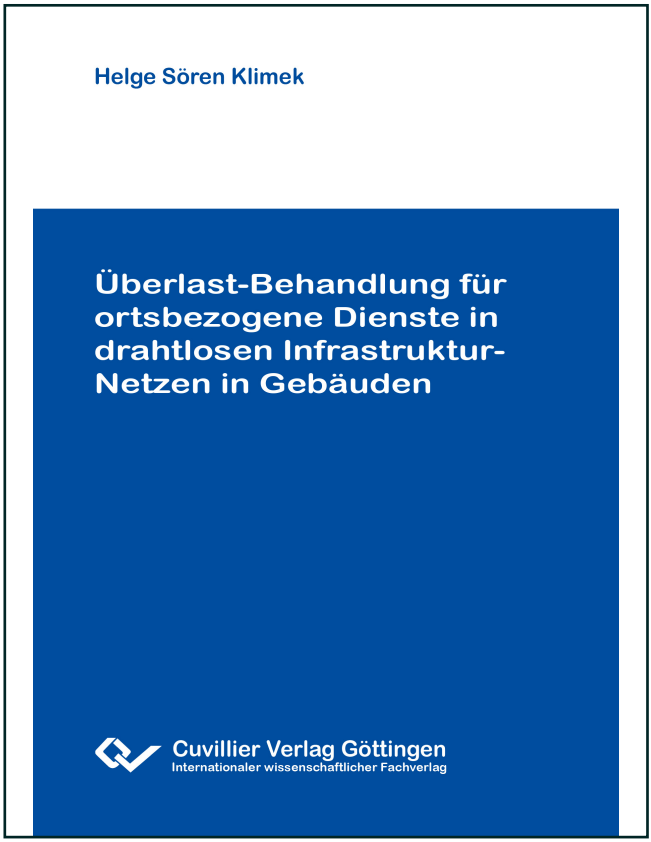
Helge Sören Klimek (Autor)
Überlast-Behandlung für ortsbezogene Dienste in drahtlosen Infrastruktur-Netzen in Gebäuden



https://cuvillier.de/de/shop/publications/6813

Copyright:
Cuvillier Verlag, Inhaberin Annette Jentzsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen, Germany
Telefon: +49 (0)551 54724-0, E-Mail: info@cuvillier.de, Website: https://cuvillier.de

Kapitel 1

Einleitung

Benutzern des Mobilfunk-Netzes ist vermutlich mindestens eine der folgenden Situationen bekannt: Zu Silvester kommen die Neujahrsgrüße per SMS mit stundenlanger Verspätung an, in der morgendlichen Rush-Hour in den S-Bahn-Tunneln ist die Internetverbindung zu langsam zum Surfen, und auf Groß-Events, zum Beispiel im Fußballstadion, kann man keinen Anruf tätigen. Diesen Dienstausfällen liegt eine Überlastung der Netz-Infrastruktur zugrunde. Beim Mobilfunk-Netz halten die Telefone Verbindung zu einer Basisstation in ihrer Nähe. Dabei teilen sich alle Mobiltelefone, die mit der selben Basisstation verbunden sind, die Ressourcen, die diese zur Verfügung stellt. Kann eine Basisstation nicht weitere Telefonate verwalten oder nicht mehr Daten transportieren, kommt es zu den beschriebenen Störungen und Ausfällen. Im Falle der Rush-Hour und der Groß-Events konzentrieren sich zu viele Nutzer auf zu wenige Basisstationen. Im Falle der Dienstausfälle zu Silvester, bei denen größere Flächen als bei Groß-Events betroffen sind, weicht das Nutzerverhalten plötzlich vom Regelverhalten ab. Während sonst nur ein Bruchteil der Nutzer ihr Mobiltelefon gleichzeitig benutzt, steigt dieser Anteil zum Jahreswechsel sprunghaft an. Die von den Basisstationen vorgehaltenen Kapazitäten sind hierauf nicht ausgelegt und in Folge dessen kommt es zu Dienstausfällen.

Grundsätzlich liegt also eine Ressourcenknappheit vor, so dass versucht werden kann, die betroffene Netzinfrastruktur auf mehr Nutzer beziehungsweise auf ein größeres zu übertragenes Datenvolumen auszulegen. Dies ist nicht immer möglich, da es Grenzen gibt, die nicht überschritten werden können, beispielsweise wenn eine verwendete Technologie bereits voll ausgebaut worden ist. Es spielen jedoch auch wirtschaftlichen Gesichtspunkte eine Rolle: Kosten entstehen, wenn Technik durch effizientere ausgetauscht wird, sowie durch Wartung und Betrieb. Ausserdem ist es für die Wirtschaftlichkeit wichtig, nicht Überkapazitäten aufzubauen, die Geld kosten, auf die aber nur in Ausnahmefällen zurückgegriffen wird.

Im Allgemeinen wird ein Netzwerk mit einem Aufbau wie dem des Mobilfunk-Netzes als drahtloses Infrastruktur-Netz bezeichnet. Es besteht aus mobilen Geräten und einem

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Backend aus fest installierten Basisstationen und Servern. Die Basisstationen und die Server sind über eine kabelgebundene Breitbandschnittstelle miteinander verbunden. Die mobilen Geräte kommunizieren per single-hop mit den Basisstationen, eine direkte Kommunikation zwischen den mobilen Geräten findet nicht statt. In diesem Zusammenhang wird der Sende- und Empfangsbereich um eine Basisstation als (Funk-) Zelle bezeichnet.

Das Mobilfunk-Netz ist dabei ein Beispiel für ein vergleichsweise leistungsstarkes Netzwerk: Mobile Nutzer können mit einer Bandbreite von vielen Megabit pro Sekunde bedient und das Netz kann von einer großen Anzahl mobiler Nutzer gleichzeitig verwendet werden. Hierzu wird ein komplexes System von hierarchisch gegliederten Basisstationen und Vermittlungsstellen mit Software und Protokollen für verschiedene Funkkanäle eingesetzt, die die zur Verfügung stehenden Ressourcen auf die Nutzer verteilen. Nicht immer sind solch leistungsfähige Netzwerke notwendig oder stehen zur Verfügung. In den letzten Jahren haben sich Funkstandards etabliert, welche eine im Vergleich zum Mobilfunk-Netz geringe Bandbreite und Reichweite aufweisen. Die Hard- und Software, die zur Kommunikation verwendet wird, ist dafür in der Regel weniger komplex und kostspielig.

Ein Beispiel hierfür ist Hardware, welche auf dem Funkstandard IEEE 802.15.4 basiert. Dessen Bandbreite ist um größenordnungen kleiner als bei Bluetooth, WLAN oder dem Mobilfunk-Netz. Der Standard zielt auf eingebettete Systeme ab, um beispielsweise Lichtschalter oder Sensoren wie Thermometer drahtlos mit anderen Systemen zu verbinden. Die Hardware ist darauf ausgelegt, wenig Energie zu verbrauchen, so dass die beschriebenen Geräte und Anwendungen über Jahre mit handelsüblichen Batterien auskommen.

Die Technologie lässt sich auch dazu einsetzten, mobile Geräte ähnlich zu Mobiltelefonen zu bauen, welche dann für spezialisierte Anwendungen zum Einsatz kommen. Ein solches Anwendungsgebiet können ortsbezogene Dienste in großen Gebäuden sein. Bei ortsbezogenen Diensten wird die aktuelle Position eines Benutzers verwendet, um einen Mehrwert zu erzeugen. Navigation, die Suche nach Lokalitäten in der Nähe oder standortabhängige Benachrichtigungen sind klassische Beispiele für solche Dienste und helfen den Benutzern, sich in unbekannten Umgebungen zurechtzufinden.

Ortsbezogene Dienste im Außenbereich sind durch Smartphones mit eingebauten Empfängern für das Globale Positionierungssystem (GPS) und durch Navigationssysteme in Fahrzeugen allgegenwärtig, aber auch in großen Gebäuden sind diese Dienste hilfreich: Am Flughafen können Passagiere den direkten Weg zum Gate, auf einem Schiff den kürzesten Weg zur eigenen Kabine finden. Auch für ortskundiges Personal können ortsbezogene Dienste einen Mehrwert bieten: Sie können die Ablaufplanung verbessern und Nutzern helfen, beispielsweise Gegenstände wie Ersatzteile oder Werkzeuge sowie Personen ausfindig zu machen.

Für ortsbezogene Dienste ist die Ermittlung genauer Positionen notwendig. Innerhalb von Gebäuden kann die Positionsermittlung jedoch nicht durch das als de-facto Standard etablierte GPS geleistet werden, da der Empfang großen Fehlern unterworfen ist oder gar nicht funktioniert und die ermittelten Positionen zu ungenau sind. Daher muss auf andere Lösungen zurückgegriffen werden, beispielsweise auf eine Erweiterung des bereits erwähnten IEEE-Funkstandards, welcher die Distanzmessung zwischen Sender und Empfänger ermöglichen soll. Mit diesen Distanzen von einem mobilen Gerät zu verschiedenen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

fest installierten Basisstationen lässt sich schließlich eine Position im Raum berechnen, welche dann die Grundlage für ortsbezogene Dienste darstellt.

Während der Verwendung von ortsbezogenen und anderen Diensten werden durch den Informationsaustausch von den mobilen Geräten mit den Basisstationen Netzwerkressourcen gebunden. Diese Ressourcen sind begrenzt und es kann dazu kommen, dass nicht mehr genügend zur Verfügung stehen, um die Kommunikation aller Teilnehmer in der Funkzelle zuverlässig aufrecht zu erhalten. Dieser Zustand wird als Überlastung bezeichnet, in dessen Folge es zu Kommunikationsabbrüchen kommt und Dienste letztlich ausfallen. Dies entspricht den am Anfang beschriebenen Ausfällen im Mobilfunk-Netz.

Die Leistungsfähigkeit der verwendeten Netztechnologie, aber auch die verwendeten Protokolle entscheiden mit darüber, wie schnell eine Überlastung eintritt: Eine grundsätzliche Schwierigkeit entsteht bei Infrastruktur-Netzen in diesem Zusammenhang aus der Verbindung von begrenzten Ressourcen der Netztechnologie und der Mobilität der Nutzer und deren unvorhersehbarem Verhalten: Benutzer können sich frei bewegen und betreten beziehungsweise verlassen dabei Funkzellen unabhängig davon, wie viele Netzressourcen bereits gebunden sind. Befinden sich zu viele Nutzer in einer Zelle und nutzen Dienste, kann dies zu einer Überlastung der Zelle führen.

Je nach Anwendungsgebiet kann eine durch Überlastung hervorgerufene Störung des Systems oder sogar ein Dienstausfall erhebliche Konsequenzen haben: Fallen ortsbezogene Dienste am Flughafen aus, welche Passagiere auf dem Weg vom Check-In bis zum Abflug unterstützen, oder Benachrichtigungsdienste, die Passagiere über Änderungen des Ablaufs informieren, ist mit erheblichen Beeinträchtigungen des Flugbetriebs und den damit verbundenen Folgen zu rechnen. Wenn Personal das System einsetzt und dieses nicht zuverlässig funktioniert, geht Arbeitszeit aufgrund fehlender Informationen oder unnötiger Wege verloren. Grundsätzlich ist davon auszugehen, dass ein System, welches auf Dauer unzuverlässig arbeitet, keine Akzeptanz finden wird.

Es stellt sich daher die Frage, ob und wie mit Überlastungen umgegangen werden kann, wenn die Netztechnologie als gegeben angenommen wird. Werden mehr Ressourcen zur Kommunikation benötigt als zur Verfügung stehen, leidet entweder die Zuverlässigkeit für alle Teilnehmer, oder es müssen Ressourcen für weniger wichtige Kommunikation aufgegeben und dafür in Kauf genommen werden, dass einige Dienste ausfallen.

Diese Arbeit beschäftigt sich mit der Frage, wie ein System zur Überlast-Behandlung bei festgesetzter Netztechnologie aufgebaut sein kann. Als Fallbeispiel werden dazu die weiter oben beispielhaft erwähnten ortsbezogenen Dienste im Flughafengebäude betrachtet. Es werden verschiedene Verfahren vorgestellt, deren Ziel die Erkennung, Reduktion und Vermeidung von Überlast ist. Die Überlast-Vermeidung nutzt anwendungsbezogenes Wissen, um Überlastungssituationen von vornherein zu umgehen. Die Erkennung bewertet den Netzzustand und bildet die Grundlage zur Steuerung der Überlast-Reduktion. Sie kann zentral oder verteilt auf den Netzknoten erfolgen und je nachdem wie das System aufgebaut wird, ist Kommunikation zur Koordinierung notwendig. Diese Kommunikation muss beachtet werden, um die Netzressourcen nicht unnötig weiter zu belasten. Die Überlast-Reduktion hat schließlich das Ziel, einer bestehenden Überlast entgegenzuwirken. In Simulationen werden die verschiedenen Verfahren einzeln untersucht und be-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

wertet. Es wird diskutiert, wie die einzelnen Teile eines Systems zur Überlast-Behandlung im Gesamtsystem angeordnet werden müssen, damit eine Überlastung nicht weiter verschärft wird. Außerdem werden die Kommunikationsprotokolle betrachtet und bewertet, in denen die Überlast-Behandlung eingebettet ist.

1.1 Struktur der Arbeit

Die Arbeit ist wie folgt strukturiert: In Kapitel 2 werden die Grundlagen und Ziele der Arbeit erläutert. Dazu werden Infrastruktur-Netze, Überlast-Behandlung und ortsbezogene Dienste definiert und das Fallbeispiel für die Arbeit beschreiben.

Kapitel 3 liefert eine detaillierte Beschreibung der Überlast-Erkennung, -Reduktion und -Vermeidung. Die Überlast-Erkennung wird dazu aufgegliedert in Ende-zu-Ende basierte Verfahren, welche sich auf Informationen an den Kommunikationsendpunkten stützt und netzunterstützte Verfahren, bei denen Informationen aus den Netzknoten bezogen werden.

Im 4. Kapitel werden beispielhaft Infrastruktur-Netze vorgestellt, darunter sind das Mobilfunk-Netz GSM, WLAN und WPAN für Sensornetzwerke. Anschließend wird die Überlast-Behandlung in anderen Netzwerken vorgestellt und beschrieben; betrachtet werden dabei das Internetprotokoll TCP/IP, ABR für ATM aus dem Bereich der Telekommunikation, WLAN im Heimgebrauch sowie verschiedene andere. Das Kapitel schließt ab mit einer Beschreibung der Fehlertoleranzmaßnahmen, wie sie im OSI-Protokollstapel Anwendung finden.

Das 5. Kapitel beschreibt das System aus Basisstationen, Server und mobilem Gerät, sowie die betrachteten Kommunikationsprotokolle. In Kapitel 6 wird das Simulations-Framework vorgestellt und die Erweiterungen zur Modellierung der Gebäudestruktur, Mobilität und des Paketaufkommens beschrieben. Das Ende des Kapitels befasst sich mit dem Modell des Flughafen Gebäudekomplexes und dem Basisstations-Layout.

Kapitel 7 führt eine allgemein gehaltene Diskussion der in Kapitel 5 vorgestellten Kommunikationsprotokolle durch. Die Betrachtung findet dabei losgelöst vom Flughafen Szenario statt. In Kapitel 8 werden die Verfahren zur Überlast-Erkennung, -Reduktion und -Vermeidung im Detail diskutiert und jeweils eine Auswahl für die weitere Betrachtung getroffen. Im 9. Kapitel werden die Simulationsszenarien beschrieben, in denen die Verfahren aus Kapitel 8 auf ihre Wirksamkeit hin untersucht werden.

In Kapitel 10 wird zunächst ein Vergleich der beiden Kommunikationsprotokolle für die vorher beschriebenen Szenarien durchgeführt. Anschließend findet in den Kapiteln 11, 12 und 13 die beschriebene Untersuchung und Bewertung der Überlast-Erkennung, -Reduktion und -Vermeidung statt.

Die Arbeit schließt in Kapitel 14 mit einer Zusammenfassung der Arbeit einschließlich der Ergebnisse und endet mit einem Ausblick auf weitere Lösungsansätze und Ideen, die sich im Verlaufe ergeben, jedoch keinen Platz mehr in der Arbeit gefunden haben.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Kapitel 2

Grundlagen und Ziele

In diesem Kapitel werden die Grundlagen für die vorliegende Arbeit erläutert. Es beginnt mit einer Beschreibung von Infrastruktur-Netzen und einer allgemeinen Erklärung von Überlast-Behandlung in Netzwerken. Anschließend wird erklärt, was unter dem Begriff ortsbezogene Dienste verstanden wird. Schließlich wird an einem Fallbeispiel erläutert, wie ortsbezogene Dienste für Passagiere im Flughafenterminal eingesetzt werden können. Zum Schluss wird für das beschriebene Fallbeispiel erklärt, welche Ziele für die Dienste mit Bezug zur Überlast-Behandlung bestehen.

2.1 Definition von Infrastruktur-Netzen

Im weiteren Verlauf der Arbeit wird von drahtlosen Infrastruktur-Netzen gesprochen; dieser Abschnitt soll daher zunächst den Begriff definieren und so ein gemeinsames Verständnis schaffen. Ein drahtloses Infrastruktur-Netz besteht aus mobilen Geräten welche drahtlos mit einem Backend aus stationären Basisstationen kommunizieren. Zum Backend gehört zusätzlich zu den Basisstationen ein Server, mit dem die Basisstationen über ein kabelgebundenes Netzwerk verbunden sind. Server und Basisstationen sind festinstallierte Systeme, welche mit dem Stromnetz verbunden sind. Ihre Energieversorgung wird deshalb als unerschöpflich angenommen. Bei dem mobilen Gerät ist die Betriebsdauer durch die Energiereserve, beispielsweise in einer Batterie oder einem Akku, begrenzt.

Der Sende- und Empfangsbereich um eine Basisstation wird (Funk-) *Zelle* genannt und ist begrenzt. Um Flächen abzudecken, die größer als eine Funkzelle sind, sind daher mehrere Basisstationen notwendig. Aufgrund dieser zellartigen Struktur wird der Begriff Zelluläres-Netz synonym zu Infrastruktur-Netz verwendet. Bewegen sich Benutzer über diese Flächen, verlassen sie die Funkzellen von Basisstationen und betreten die Zellen anderer Stationen. Üblicherweise wird versucht, im Backend Informationen darüber zu erlangen, in welcher Funkzelle sich ein mobiles Gerät befindet. Diese Ortsinformation wird dazu ge-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

nutzt um Daten direkt über die Basisstation zu senden, in deren Funkzelle sich ein Gerät befindet.

Die mobilen Geräte empfangen und senden Daten immer single-hop zu den Basisstationen, eine direkte Kommunikation zwischen den mobilen Geräten ist nicht vorgesehen. Sollen Nachrichten zwischen zwei mobilen Geräten ausgetauscht werden, erfolgt die Kommunikation immer über mindestens eine Basisstation. Die Kommunikationsrichtung von einer Basisstation zum mobilen Gerät wird als *Downlink* bezeichnet, die Richtung vom mobilen Gerät zur Basisstation als *Uplink*.

Der Begriff Infrastruktur-Netz wird in IEEE 802.11 WLAN verwendet und bezeichnet dort, dass mobile Geräte Daten über eine Basisstation austauschen. Durch diesen Aufbau müssen alle mobilen Geräte lediglich zur Basisstation Funkkontakt halten können, nicht jedoch zu jedem anderen mobilen Gerät. Die Basisstation dient zusätzlich dazu, die mobilen Geräte mit einem verkabelten Netzwerk zu verbinden. Mit Hilfe des Extended Service Sets (ESS) kann auch mit WLAN-Basisstationen ein zelluläres Netzwerk aufgespannt werden, bei dem mobile Geräte beim Zellwechsel den Kontakt zum Netz halten. Informationen über Zellwechsel werden den Basisstationen über das verkabelte Netzwerk mitgeteilt und Nachrichten über die Basisstation gesendet, an der das mobile Gerät registriert ist.

Eines der größten und bekanntesten Zellulären-Netze ist das GSM Mobilefunk-Netzwerk, bei dem Funkzellen zu Location Areas (LA) zusammengefasst und mehrere Zellen von einem Basestation-Controller gesteuert werden.

2.2 Überlast-Behandlung in Netzwerken

In der Literatur werden die Begriffe Überlast-Vermeidung (Congestion-Avoidance) und Überlast-Reduktion (auch Überlast-Kontrolle, engl. Congestion-Control) teilweise uneinheitlich verwendet. Für den weiteren Verlauf der Arbeit bezeichnet Überlast-Vermeidung Verfahren, die darauf abzielen, eine Überlastungsituation gar nicht erst entstehen zu lassen. Das Ziel der Überlast-Reduktion ist das möglichst zügige Auflösen von Überlastsituation. Beide Verfahren sind Mechanismen der Überlast-Behandlung. Eine Erklärung zu Überlast-Reduktion kann in [KR08] gefunden werden; die folgende Erklärung baut darauf auf. In Kapitel 3 wird ausführlicher auf das Thema Überlast-Behandlung eingegangen.

Computer-Netzwerke, ganz gleich ob verkabelt oder drahtlos, sind aus Komponenten aufgebaut, die begrenzte Ressourcen verwenden und zum Betrieb benötigen. Sind diese Ressourcen gebunden, bricht die Funktionalität des Netzwerks zusammen. Beispiele für diese Ressourcen sind das Medium, über das Daten übertragen werden, oder Pufferspeicher in den Netzwerkknoten.

Je nach verwendeter Technologie kann das Medium nur von einem Teilnehmer zur Zeit verwendet werden um Daten zu senden. Das Übertragen dauert zudem eine bestimmte Zeit, in der das Medium dann belegt ist. Sendet in dieser Zeit ein anderer Teilnehmer, kommt es zu Übertragungsfehlern. Puffer in Netzwerkknoten auf dem Pfad von Quelle zum Ziel sind durch ihre Größe begrenzt und werden dazu verwendet, Daten vor dem Weiterleiten zwischenzuspeichern. Werden Daten schneller empfangen als sie verschickt

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

werden können, laufen die Puffer voll. Sind sie schließlich voll, müssen Pakete verworfen werden, und es kommt zum Paketverlust. Während sich die Pufferstände in einem Netzwerkknoten aufbauen, nimmt die Verzögerung der Zustellung zu, da Pakete erst einmal im Puffer warten müssen, ehe sie weitergeleitet werden können.

Um die Übertragungszuverlässigkeit zu steigern, wird bei Paketverlusten normalerweise eine erneute Übertragung des Pakets ausgelöst. Dieser Vorgang allein behandelt allerdings nur das Symptom, also den Paketverlust, nicht jedoch die Ursache des Verlusts, also eine mögliche Überlastung des Netzes. So können Neu-Übertragungen von Paketen den Zustand des Netzes weiter verschlechtern. Nach einer begrenzten Anzahl von versuchten Neu-Übertragungen bricht die Kommunikation schließlich ab.

Um Überlastung zu vermeiden werden Informationen benötigt, die Aufschluss über den Zustand des Netzes geben, zum Beispiel über Pufferstände in den Netzknoten oder die Auslastung des Mediums. Liefern Netzwerkkomponenten direkt Feedback, wird von Netzwerkunterstützter Überlast-Reduktion (network assisted congestion control [sic!]) gesprochen. Es existieren unterschiedliche Ansätze; zum Beispiel können Netzwerkknoten bei passierenden Paketen ein Überlast-Bit setzen, die Datenrate vorschreiben oder sogar Nachrichten als Warnung vor Überlast (auch: *Choke Message*) verschicken.

Es kann auch versucht werden, anhand des Netzwerkverhaltens auf dessen Zustand zu schließen. Ist ein Netz normalerweise sehr zuverlässig, können Paketverluste als Indikator verwendet werden oder eine zunehmende Dauer der paketübermittlung. Da hier die beiden Kommunikationspartner Informationen austauschen müssen, wird auch von einer Ende-zu-Ende Überlast-Behandlung gesprochen. Je nach Art des Datenverkehrs und Anzahl der Nutzer kann sich die Belastung der Netzknoten und Verbindungen jedoch ständig ändern, wenn Verbindungen von Teilnehmern auf- oder abgebaut werden.

Zusätzlich können Kosten in Betracht gezogen werden, die entstehen, wenn ein Paket entlang eines Pfads verworfen wird. Wird durch einen vollen Puffer ein Paket verworfen, dann ist die Übertragungskapazität, die bisher zum Weiterleiten verwendet wurde, umsonst gewesen. Es ist günstiger ein Paket zu verwerfen, welches möglichst wenige Netzwerkknoten passiert hat. Auch wenn die Verzögerung von Paketen in überlasteten Netzwerkknoten eine festgelegte Grenze überschreitet, nimmt ein Sender an, dass das Paket verloren gegangen ist, und löst eine Neu-Übertragung aus. In diesem Fall werden Pakete erneut gesendet, obwohl die gleichen Pakete noch in einem Puffer des Knotens auf Zustellung warten.

2.3 Beschreibung von ortsbezogenen Diensten

Im folgenden wird erläutert worum es sich bei ortsbezogenen Diensten handelt, und diese werden inhaltlich und historisch eingeordnet. Anwendungsbeispiele zeigen, wo diese Dienste in Verwendung sind. Abschließend werden zwei Klassifizierungen für ortsbezogene Dienste gegeben, welche sich auf die Kommunikations- und Interaktionsmuster der Dienste stützen. Die folgende Beschreibung baut auf [KÖ5], [SV04] und [KH06] auf.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Ortsbezogene Dienste (location-based services, LBS) oder auch ortsbewusste Dienste (location-aware services) sind eine Spezialform von kontextbewussten Diensten (context-aware services). Diese Dienste passen ihr Verhalten an Parameter an, welche den Kontext eines Ziels reflektieren. Durch die Benutzung der Position mindestens eines Ziels soll der Dienst einen Mehrwert erzielen. Das Ziel, dessen Position bestimmt wird, muss dabei aber nicht zwangsläufig der Benutzer des Dienstes sein.

Für ortsbezogene Dienste ist die Kenntnis einer Position notwendig, entweder der eigenen Position oder der eines Objekts oder einer anderen Person. Zur Positionsbestimmung außerhalb von Gebäuden hat sich in den letzten Jahrzehnten das US amerikanische Globale Positionierungssystem (GPS) als de-facto Standard etabliert. Dieser Verbreitung ging 1980 die Entscheidung der US-Regierung voraus, das System für nicht militärische Anwendungen zu öffnen. Seit dem wurde GPS zur Flugverkehrssteuerung, Hafensteuerung, zum Flottenmanagement und in Fahrzeug-Navigationssystemen eingesetzt.

1996 hat die US amerikanische Federal Communications Commission (FCC) Mobilfunkbetreiber dazu verpflichtet, bei Notrufen aus den Mobilfunk-Netzen die Position der Anrufer automatisch zu erfassen und diese an die Notrufzentralen weiterzuleiten. Diese Verpflichtung trägt der Tatsache Rechnung, dass es immer mehr Notrufe von Mobiltelefonen gibt, bei denen von der Telefonnummer nicht mehr auf eine Adresse geschlossen werden kann. Zusätzlich sind viele der Anrufer nicht in der Lage, ihre Position genau zu beschreiben. Die Genauigkeit, mit der die Position ermittelt und an die Notrufzentralen übermittelt werden muss, wurde auf zwei Phasen verteilt und festgeschrieben. Für Europa bestehen seit 2002 ähnliche Bestrebungen, jedoch ohne die Festlegung von Ortungsgenauigkeiten.

Neben der reinen Ortung sind in der Regel weitere Systeme beteiligt, um ortsbezogene Dienste zu erbringen. Oft ist die Kenntnis über eine Position, zum Beispiel in Breiten- und Längengraden, allein nicht hilfreich. Stattdessen werden ortsbezogene Informationen wie Straßennamen, Hausnummern, Bebauung oder anderes benötigt. Hierzu werden Geographische Informationssysteme (GIS) verwendet, mit denen zu einer Position ortsbezogene Information ermittelt und abgelegt werden können.

Insbesondere durch die weite Verbreitung von GPS-fähigen Smartphones in den letzten Jahren haben ortsbezogene Dienste in viele Bereichen Einzug in den Massenmarkt gefunden. Die wohl am weitesten verbreiteten und bekanntesten sind ortsbezogene Informationsabfragen für Points of Interest (POI), welche üblicherweise nach Nähe und Relevanz sortiert und gefiltert sind. Ein Beispiel hierfür ist Google Places[1] auf Android Smartphones, welche es erlaubt den nächsten Geldautomaten, Hotel, Restaurant oder Cafe zu finden.

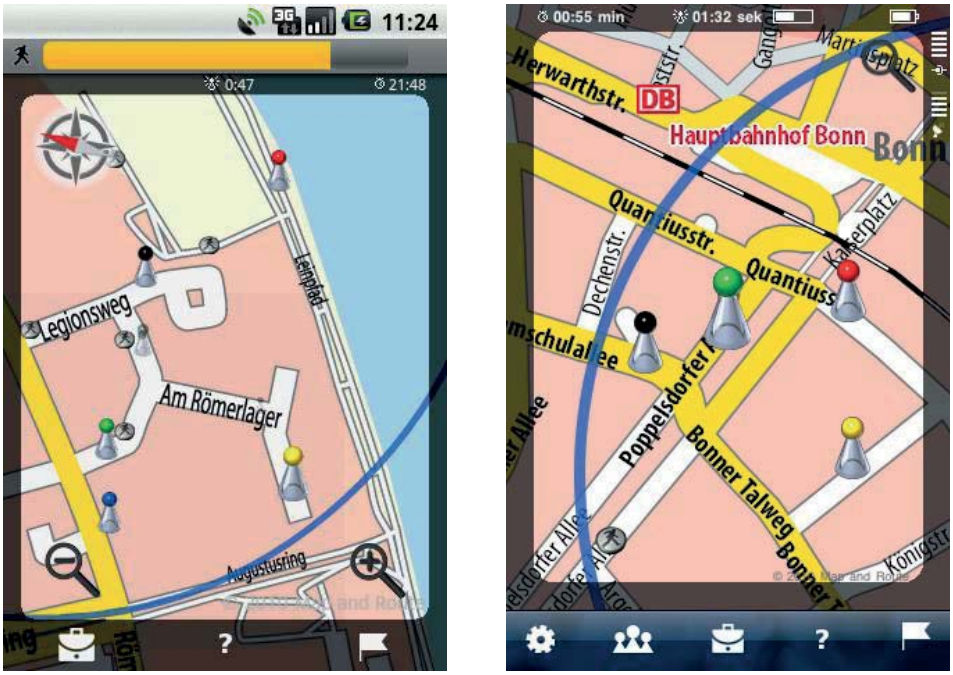
Komplexer sind Dienste, welche aktiviert werden, wenn sich der Benutzer einer bestimmten Position nähert oder diese verlässt (auch: engl. geofencing), sowie sogenannte Community Services. Bei Community Services wird die eigene Position anderen zur Verfügung gestellt, so dass diese anderen auf einer Karte dargestellt werden kann. Üblicherweise kann festgelegt werden, wer die eigene Position sehen kann und mit welcher

[1] http://www.google.com/places

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Genauigkeit. Google Latitude[2] ist ein Beispiel für einen solchen Dienst. Bei Apples iPhone können Erinnerungen mit dem Betreten oder Verlassen eines Orts verknüpft werden[3]. In diesem Fall verlässt die ermittelte Position das Smartphone nicht; es muss jedoch regelmässig geprüft werden, ob die angegebene Position erreicht oder verlassen wurde.

Auch in der Verkehrstelematik werden ortsbezogene Dienste verwendet: Sie informieren Kraftfahrer über Baustellen, Staus, Wetter, vorhandene Parkplätze oder werden für die Mauterhebung eingesetzt. Bei der Flotten-Planung helfen sie bei der Errechnung der Zustellzeit oder dabei, neue Aufträge an Fahrzeuge in der Nähe zu delegieren.



(a) Mr. X Mobile - Android

(b) Mr. X Mobile - iTunes

Abbildung 2.1: Screenshots der ortsbezogenen Spiele Mr. X Mobile, Android (2.1(a)) und iTunes (2.1(b)).

Ortsbezogene Spiele sind ein weiteres Anwendungsgebiet für ortsbezogene Dienste; hier wird die virtuelle Spielwelt mit der geographischen Beschaffenheit der realen Welt vermischt. Das für Android[4] und iOS[5] verfügbare Spiel Mr. X Mobile (Abbildung 2.1) ist dem Ravensburger Brettspiel „Scotland Yard“ beziehungsweise „Mister X, Flucht durch Europa“ nachempfunden. Im Spielverlauf versteckt sich eine Person, Mr. X, in einer Stadt, während die anderen Spieler versuchen ihn aufzuspüren. Dabei wird die Position von Mr. X zu unterschiedlichen Zeitpunkten und Anlässen den anderen Spielern auf ihren Smartphones enthüllt.

[2] http://www.google.com/latitude
[3] http://www.apple.com/de/iphone/features/siri-faq.html
[4] http://market.android.com/details?id=edu.bonn.cs.amoga.mrx.android
[5] http://itunes.apple.com/de/app/mister-x-mobile/id373751016

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Ortsbezogene Dienste werden in der Literatur unterschiedlich fachlich und technisch klassifiziert; bei den folgenden Definitionen liegt der Fokus auf den Kommunikations- und Interaktionsmustern. Eine Klassifikation kann danach erfolgen, wer geortet wird. Ein personenorientierter oder *aktiver Dienst* wird vom Benutzer kontrolliert; das Ziel besteht darin, den Benutzer selbst zu orten und mit den Ortsinformationen einen Mehrwert für den Nutzer zu erzielen. Ein typisches Beispiel ist die ortsbezogene Informationsabfrage. Dem gegenüber steht der geräteorientierte oder *passive Dienst*, bei dem der Benutzer nicht notwendigerweise das Ziel der Ortung ist; dies könnte ebensogut eine andere Person oder ein beliebiges Objekt sein. Die geortete Person oder das Objekt kontrolliert den Dienst also nicht. Beispiele hierfür sind Personen- oder Objekt-Verfolgung wie beim Flotten-Management oder für den Diebstahlschutz.

Eine andere Klassifikation kann nach der Aktivierung der Dienste erfolgen: *Pull* oder *reaktive Dienste* werden aktiv durch den Nutzer gestartet und sind charakterisiert durch ein synchrones Interaktionsmuster: Der Dienst wird aufgerufen, es werden Ortsinformationen gesammelt und verarbeitet und schließlich ein Ergebnis zurückgegeben. Bei *Push* oder *proaktiven Diensten* empfängt der Nutzer Daten, ohne diese explizit anzufragen. Die Aktivierung erfolgt automatisch, wenn ein zuvor definiertes Ortsereignis (Location Event) eintritt, beispielsweise das Betreten oder Verlassen eines bestimmten Gebiets oder einer Position. Das Interaktionsmuster ist asynchron und die Ortung findet permanent statt. Diese Dienste benötigen mehr Ressourcen als Pull-Dienste, da die Position regelmässig erfasst und verarbeitet werden muss. Neben der Positionsbestimmung ist so möglicherweise auch noch Kommunikation zur Diensterbringung notwendig.

2.4 Fallbeispiel: Ortsbezogene Dienste im Flughafenterminal

Ortsbezogene Dienste verwenden die aktuelle Position zum Beispiel des Benutzers, um bei der Diensterbringung einen Mehrwert für den Benutzer zu erbringen. Typische Anwendungen sind die Suche nach Lokalitäten in der Umgebung oder die Navigation zu einem ausgewählten Ziel. Sie können auch in großen Gebäuden mit vielen Nutzern, wie Bahnhöfen, Krankenhäusern oder Messen, die sich über sehr große Flächen mit großen Räumen und über mehrere Stockwerken erstrecken, angeboten werden.

Um ortsbezogene Dienste verwenden zu können, ist ein mobiles Gerät notwendig, welches zum einen eine Ortung im Gebäude ermöglicht und zum anderen Zugang zu den Diensten erlaubt. Dies beinhaltet auch die Möglichkeit, mit den entsprechenden Systemen zum Beispiel über eine bestehende Infrastruktur kommunizieren zu können. Für die Ortung hat sich außerhalb von Gebäuden GPS etabliert. Innerhalb von Gebäuden ist der Empfang von GPS jedoch großen Fehlern unterworfen, so dass die ermittelten Positionen ungenau werden – oder eine Positionsbestimmung gar nicht mehr möglich ist. Um ortsbezogene Dienste in Gebäuden anbieten zu können, ist deshalb ein System zur Positionierung notwendig. Dieses kann zum Beispiel in ein drahtloses Infrastruktur-Netz integriert werden.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Das in dieser Arbeit untersuchte Fallbeispiel wird aus dem Projekt Airport 2030 abgeleitet. In dem Projekt wird untersucht, wie Passagierprozesse im Flughafenterminal anhand eines kostengünstigen, mobilen Geräts optimiert werden können. Bei dem Gerät handelt es sich um die Digitale Boarding Assistenz (DigiBA) welche die Boardkarte ersetzen soll. Sie wird Passagieren gegeben um diese beim Durchlaufen der Flughafenprozesse, vom Check-In bis zum Boarding, zu unterstützen. Die DigiBA besitzt eine Ortungstechnologie und kann Daten über die Infrastruktur mit einem Server austauschen. Die angebotenen Dienste decken dabei sowohl Pull- als auch Push-Dienste ab und stellen dabei jeweils andere Anforderungen an die Netzinfrastruktur:

- **Indoor-Navigation:** Benutzer haben die Möglichkeit, die Zielführung innerhalb des Gebäudes von der aktuellen Position zu einem ausgewählten Ziel zu nutzen. Die Ortung findet dabei auf dem mobilen Gerät statt, von der konkreten Ortungstechnologie wird hier jedoch abstrahiert. Die Position wird an den Server geschickt, welcher die Routenberechnung vornimmt und Navigationsanweisungen an das Gerät zurücksendet. Um den Dienst auszuführen ist es notwendig, dass über den Zeitraum des Navigationsprozesses regelmässig Nachrichten von und zum mobilen Gerät transportiert werden können. Zusätzlich haben die einzelnen Positions-Updates und Navigationsanweisungen einen hohen zeitlichen Bezug und müssen daher mit festgelegter maximaler Verzögerung zugestellt werden. Es sei angenommen, dass Navigationsanweisungen mindestens alle 30 m angezeigt werden müssen. Als mittlere Gehgeschwindigkeit seien 1.34 m/s angenommen. Müssen Positionsdaten zunächst vom mobilen Gerät zum Server gesendet werden, ergibt sich eine Server-Roundtrip-Time (RTT) von rund 22 s. Wenn die Verarbeitungszeit idealisiert mit 0 s angenommen wird und eine gleiche Übertragungszeit in beide Kommunikationsrichtungen, von mobilen Gerät zum Server und zurück, so ergibt sich eine maximale Nachrichtenlatenz von 11 s pro Kommunikationsrichtung.

- **Ortsbezogene Benachrichtigung:** Passagiere sollen benachrichtigt werden, wenn Sie einen bestimmten Bereich betreten oder diesen verlassen. Hierzu muss das mobile Gerät relevante Änderungen seiner Position dem Server mitteilen, welcher gegebenenfalls mit einer Nachricht an das mobile Gerät reagiert. Diese Benachrichtigung kann zum Beispiel verwendet werden, um Passagieren im Check-In-Bereich mitzuteilen, welche Dokumente sie für das Check-In bereithalten müssen oder um vor den Schaufenstern von Geschäften ortsbezogene Werbebotschaften zu versenden.

- **Informationsabfrage:** Vom mobilen Gerät aus werden Informationen vom Server abgefragt. Das mobile Gerät sendet eine Nachricht an den Server und dieser antwortet mit einer oder mehreren Nachrichten. Informationsabfragen dienen allgemein dazu, dem Passagier die Möglichkeit zu geben, weitere Informationen einzuholen. Ein Beispiel hierfür ist das Erfragen des Wetters am Zielort. Dieser Anwendungsfall ist ortsbezogen, wenn für die Erbringung des Dienstes die Position des

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

mobilen Geräts notwendig ist, zum Beispiel bei der Suche nach dem nächstgelegenen Wickelraum. In diesem Fall muss für die Anfrage zunächst die Position ermittelt und mit übertragen werden.

- **Benachrichtigung:** Es können unaufgefordert Nachrichten an ein mobiles Gerät gesendet werden. Dieser Anwendungsfall ist nicht ortsbezogen, auch wenn zum Zustellen der Nachrichten die Kenntnis über die Funkzelle, in der sich das mobile Gerät befindet, von Vorteil wäre. Diese Benachrichtigungsform ist grundsätzlich notwendig, um Passagiere über den weiteren Verlauf des Boardingprozesses, zum Beispiel bei Änderungen des Gates oder der Abflugzeit, zu informieren oder um Passagiere zum Boarding aufzurufen – für solche Fälle kann der Erhalt der Nachricht optional vom Benutzer auch noch bestätigt werden.

2.5 Ziele für das beschriebene Fallbeispiel

In dem zuvor beschriebenen Fallbeispiel können ungewöhnliche Belastungen der Infrastruktur auftreten. Ein Beispiel dafür ist, wenn sich auf einmal viele Benutzer in einen kleinen Bereich eines Gebäudes, in eine Funkzelle, bewegen. Allein durch Steuerungsdaten, die zwischen dem mobilen Gerät und der Infrastruktur ausgetauscht werden, kann es hier zum Beispiel durch viele Anmeldevorgänge zu einer großen Belastung des Funkkanals und der Infrastruktur kommen. Auf der anderen Seite kann auch eine untypische Belastung der Infrastruktur entstehen, wenn auf einmal wenige Benutzer in einer Funkzelle ein sehr großes Datenaufkommen erzeugen. Dieser Fall könnte beispielsweise eintreten, wenn das Abfluggate für ein Flugzeug verlegt wird und sehr viele Passagiere die Indoor-Navigation starten. Bei beiden Beispielen kommt die Mobilität der Benutzer erschwerend hinzu, da sie ihnen jederzeit erlaubt, von einer Funkzelle in eine andere weiterzugehen und so eine ohnehin belastete Funkzelle weiter zu belasten.

Fehlertoleranz beschreibt die Eigenschaft eines Systems, seine Funktionalität auch unter unvorhergesehenen Ereignissen aufrechtzuerhalten, so dass Dienste entsprechend ihrer Spezifikation ausgeführt werden können. Hierzu müssen die Grenzen des beschriebenen Systems bekannt sein. Die Grenzen für das in dieser Arbeit betrachtete Infrastruktur-Netz sind jedoch nicht bekannt. Es fehlen Erfahrungswerte, da Infrastruktur-Netze normalerweise mit leistungsfähigeren Technologien wie WLAN, GSM, UMTS oder LTE aufgebaut werden. Hinzu kommt, dass durch unterschiedliche Anwendungen, unterschiedliche Protokolle und die Mobilität der Benutzer die Abschätzung weiter erschwert wird. Ziel der Arbeit ist es deshalb, in einem ersten Schritt die Kapazitätsgrenze des Systems zu ermitteln. Hierzu wird das zuvor beschriebene Fallbeispiel in einem Netzwerksimulator abgebildet und die Belastungsgrenzen ermittelt.

Im Falle einer großen Belastung des Netzwerks wird davon ausgegangen, dass sich die Qualität der Funkverbindungen mit zunehmender Anzahl der Daten-Pakete weiter verschlechtert. Je mehr Pakete ausgetauscht werden müssen, desto mehr Kollisionen treten auf, und desto mehr Pakete gehen verloren. Verlorengegangene Pakete lösen erneute Übertragungen aus, welche sich weiter negativ auf die Verbindungsqualität auswirken.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Zur Behandlung müssen die beteiligten Geräte feststellen, dass der Funkkanal in einer Zelle an der Kapazitätsgrenze angelangt ist, und versuchen die Kommunikation zu drosseln. Dabei muss vermieden werden, die Kommunikation abrupt abzubrechen, da dies zu einer Dienstunterbrechung der gerade laufenden Anwendungen führen würde.

Durch die Mobilität der Benutzer ist es zudem möglich, dass Nutzer aus einer wenig frequentierten Funkzelle in eine Zelle wechseln, die an der Leistungsgrenze arbeitet. So tragen die Nutzer zunächst sogar mit ihren laufenden Übertragungen weiter zur Belastung der Zelle bei. Generell kann versucht werden, steuernd auf die Infrastruktur einzuwirken. In vielen Fällen ist hierfür jedoch auch die Kommunikation mit den mobilen Geräten notwendig. Die dazu benötigten Nachrichten können jedoch nur zuverlässig zugestellt werden, wenn die Kapazitätsgrenze noch nicht erreicht ist. Aufgrund der mobilen Nutzer scheint daher die Erstellung einer globalen Sicht auf das Netzwerk und seines Zustands sinnvoll. Notwendig ist es zudem, frühzeitig zu erkennen, wann und wo Überlastungen entstehen, so dass gegengesteuert werden kann solange die notwendige Kapazität noch verfügbar ist.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Kapitel 3

Maßnahmen zur Überlast-Behandlung

Auf der einen Seite stellen verteilte Anwendungen Anforderungen an Netzwerke um ihre Dienste mit der von ihnen geforderten Dienstgüte zu erbringen. Netzwerke, auf der anderen Seite, sind begrenzte Ressourcen, bei denen Betriebsmittel während der Benutzung gebunden werden. Werden von den Anwendungen Anforderungen gestellt, die so viele Netzwerkressourcen binden, dass die Anwendungen deshalb nicht mehr ihre Gütekriterien erfüllen können, wird im Folgenden von *Überlast* gesprochen. In solch einem Fall können Anwendungen ihre Dienste nicht mehr mit der von ihnen geforderten Güte erbringen. Als Belastung des Netzwerks (kurz: *Netzlast*) werden dabei die von den Anwendungen gestellten Anforderungen an das Netzwerk bezeichnet.

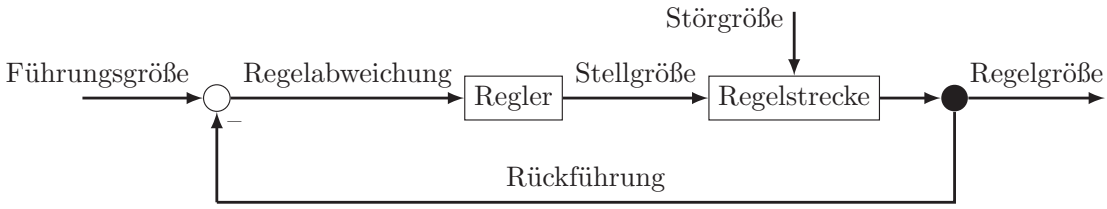


Abbildung 3.1: In einem Regelkreis wird eine Störgröße über einen Regler ausgeregelt, so dass die Regelabweichung zwischen Führungsgröße und Sollgröße möglichst gering ist.

Überlast-Erkennung und Überlast-Reduktion arbeiten in einem *Regelkreis* (Abbildung 3.1) zusammen, um das Eintreten von Überlast zu vermeiden oder, falls diese doch eingetreten ist, zügig aufzulösen. Die Überlast-Erkennung ermittelt hierzu die aktuelle Netzlast als Regelgröße. Diese wird mit einem zuvor festgelegten Sollwert verglichen. Wird eine Regelabweichung ermittelt, liegt eine Überlast vor und muss ausgeregelt werden. Als Regler dient die Überlast-Reduktion, welche die Netzlast verringert. Die Reduktion erfolgt solange bis die Sollgröße, also die Netzlast, wieder unterschritten wird.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Der Regelkreis aus Überlast-Erkennung und Überlast-Reduktion stellt ein *reaktiv*es System dar, das sich auf eine Lastsituation einstellt und ihr entgegenwirkt. Da die Grenze zwischen Last und Überlast fließend verläuft, soll in dieser Arbeit *Überlast-Vermeidung* als Verfahren abgegrenzt werden, welche *proaktiv* eine Überlast vor deren Entstehen zu verhindern versucht.

Überlast-Erkennung, -Reduktion und -Vermeidung sind Teile der Überlast-Behandlung (siehe Abbildung 3.2). Kann eine Überlast-Behandlung nicht erfolgreich durchgeführt werden, wird der Zustand des Systems unsicher. Laufende Anwendungen können ihre Dienste nicht mehr mit der geforderten Güte erbringen und liefern eventuell falsche Ergebnisse oder fallen ganz aus.

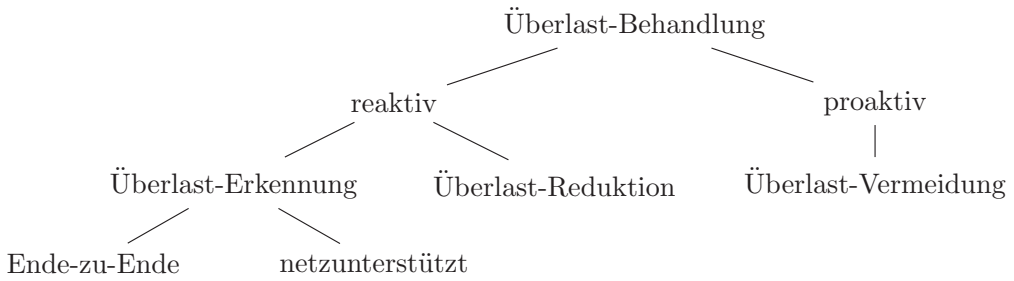


Abbildung 3.2: Überlast-Behandlung setzt sich aus drei Teilen zusammen: Erkennung und Reduktion reagieren auf eintretende Überlast und Überlast-Vermeidung versucht ein Eintreten proaktiv zu verhindern.

Überlast-Erkennung und -Reduktion können räumlich und zeitlich voneinander getrennt durchgeführt werden. Diese zeitliche und räumliche Trennung muss überwunden werden, damit beide zusammenarbeiten können. Wo eine Erkennung durchgeführt werden kann, hängt davon ab, wo die notwendigen Informationen ermittelt werden, und somit vom Erkennungs-Verfahren. Sind Erkennung und Reduktion räumlich getrennt, wird es notwendig, die Überlast-Informationen zum Ort der Reduktion zu übertragen. Hier müssen unter Umständen Parameter wie Zuverlässigkeit und Latenz der Übertragung sowie der erzeugte Overhead berücksichtigt werden, damit die Reduktion effektiv durchgeführt werden kann.

Bei der Übertragung, insbesondere bei Multi-Hop-Netzwerken und Überlast, entsteht eine zeitliche Verzögerung. Durch diese verzögert sich auch der Beginn der Reduktion, so dass sich die Zeitspanne zwischen Erkennung und Reduktion vergrößert. Zusätzliche Pakete für die Übertragung der ermittelten Überlast-Daten schmälern zudem die Nettobandbreite für Anwendungs-Daten. Werden die Überlast-Daten erst im Falle von eintretender Überlast gesendet, vergrößern sie diese möglicherweise noch. Mit Bezug zu dem im Fallbeispiel (siehe Kapitel 2.4) verwendeten Infrastruktur-Netz ergeben sich unterschiedliche Möglichkeiten, die Überlast-Erkennung und -Reduktion auf mobile Geräte, Basisstationen und Server zu verteilen. Im Folgenden werden drei Möglichkeiten vorgestellt:

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.