Fachhochschule Wedel
University of Applied Sciences

Seminar WS 2004/2005

Mobile Computing

Thema: Mobile Transportschicht

Gunnar Adler (mi7933)

1. Einleitung

Anwendungen nutzen zur Kommunikation in Netzen die Protokolle der Transportschicht. Für Anwendungen soll es dabei unerheblich sein, welche Art von Übertragungsmedium genutzt wird.
Die wichtigsten Protokolle der Transportschicht sind TCP und UDP, die auf dem Protokoll IP der Vermittlungsschicht aufbauen. Die eindeutige Adressierung der Anwendungen erfolgt durch Ports.

Damit Anwendungen bei der Kommunikation in drahtlosen Netzen die gleichen Rahmenbedingungen annehmen können wie in drahtgebundenen Netzen, ist neben der grundlegenden Unterstützung von Mobiltät auf Ebene der Vermittlungsschicht durch Mobile IP auch eine Anpassung der Protokolle der Transportschicht notwendig.

Da UDP bis auf die Adressierung mit Ports gegenüber IP über keine weiteren Mechanismen verfügt, ist hier allerdings keine Anpassung nötig und möglich.
So bietet UDP im Gegensatz zu TCP keine Garantien über die Auslieferung oder die richtige Reihenfolge von Daten zwischen Anwendungen. Treten Paketverluste auf, muß die Anwendung selbst darauf reagieren können.

TCP besitzt einige besondere Eigenschaften, die in drahtgebundenen Netzen eine reibungslose Kommunikation ermöglichen aber in drahtlosen Netzen die Leistungsfähigkeit stark vermindern kann.
TCP wurde für drahtgebundene Netze entwickelt und ist nicht an die Besonderheiten der mobilen Kommunikation angepaßt.

Ein wichtiger Mechanismus von TCP ist die *Staukontrolle*. TCP nimmt bei Paketverlust einen Stau an einem Punkt der Übertragungsstrecke an und drosselt die Senderate. UDP würde mit der gleichen Rate weitersenden und so der Auflösung des Staus entgegenwirken.

In Festnetzen mit festen Endsystemen treten kaum Übertragungsfehler auf, deswegen ist der Grund für Paketverlust fast immer ein Stau an einem Netzknoten wie einem Router.
Ein Stau ist die temporäre Überlastung an einem Punkt der der Übertragungsstrecke. Durch hohes Verkehrsaufkommen füllt sich der Paketpuffer eines Routers, da die Kapazität eines Ausgangs zu klein ist. Die einzige Möglichkeit des Routers darauf zu reagieren ist, einige Pakete nicht weiterzuleiten und zu verwerfen.
Dem Empfänger macht sich dann eine Lücke im Datenstrom bemerkbar.

Da es bei TCP keine explizite Signalisierung von fehlenden Paketen gibt, bestätigt der Empfänger weiterhin alle lückenlos bis zum fehlenden Paket empfangenen Pakete. Dadurch nimmt der Sender eine Überlastung auf der Übertragungsstrecke an und reduziert die Senderate. Eine Wiederholung und fortgesetztes Senden mit gleicher Senderate ist nicht sinnvoll, da der Stau nicht abgebaut wird und es weiterhin zu Paketverlusten kommt.

Alle anderen TCP-Verbindungen, die einen Stau durch Paketverluste bemerken, reagieren in gleicher Weise. So kann ein Stau schnell abgebaut werden und die Aufteilung der Bandbreite unter allen Nutzern wird garantiert. Dies trägt maßgeblich zur Stabilität des Internets bei.

Das Verhalten nach der Erkennung eines Staus wird als *Slow Start* bezeichnet. Der Sender berechnet hierzu fortlaufend ein Staufenster. Die Größe des Staufensters beträgt anfangs ein Segment bzw. Paket und wird danach bei Eintreffen von Bestätigungen verdoppelt. Da durch dieses expotentielle Wachstum die Verdoppelungsschritte irgendwann so groß werden, daß es sehr wahrscheinlich ist, unmittelbar einen Stau zu verursachen wird an einem Stauschwellenwert zu linearem Wachstum übergegangen. Dieser Stauschwellenwert wird auch als Congestion Treshold bezeichnet.
Bei ausgebliebenen oder mehrfach für das gleiche Paket empfangenen Bestätigungen wird der Stauschwellenwert auf die Hälfte des aktuellen Staufensters und das Staufenster auf die Größe 1 gesetzt. So beginnt der Slow Start Mechanismus mit einem angepaßten Stauschwellenwert erneut.

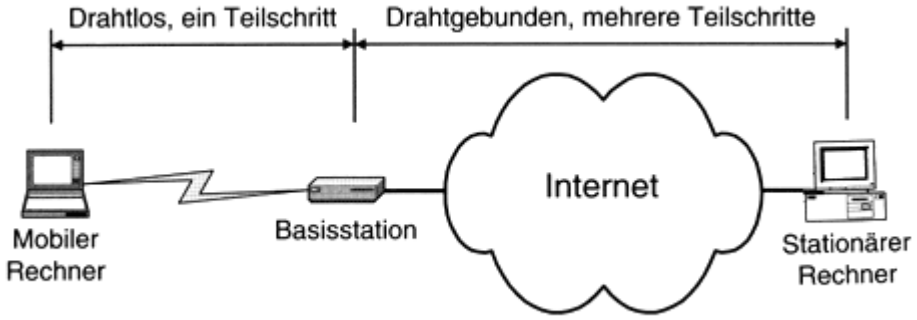
Ein weiterer Mechanismus, um auf Paketverluste zu reagieren ist *Fast Retransmit/Fast Recovery*. Hierbei wird aus dem weiteren Empfang von Bestätigungen auf eine nur kurzfristige Stausituation oder einen Übertragungsfehler geschlossen.
Der Einsatz von Slow Start wäre hier nicht sinnvoll, da kein schwerer Stau vorliegt. Stattdessen überträgt der Sender die fehlenden Pakete sofort erneut (Fast Retransmit) und fährt ohne Verkleinerung des Staufensters mit der hohen Senderate fort (Fast Recovery).
Auf diese Weise kann die Leistungsfähigkeit von TCP sehr gesteigert werden. Wenn Bestätigungen allerdings gänzlich ausbleiben, muß weiterhin Slow Start eingesetzt werden, da von einem schweren Stau ausgegangen wird.

Der Einsatz eines klassischen TCP mit Slow Start kann die Leistungsfähigkeit von mobilen Geräten stark verschlechtern, da aus dem Ausbleiben von Bestätigungen oft fälschlicherweise auf einen Stau im Netz geschlossen wird. In drahtlosen Netzen sind die Hauptursachen für Paketverlust viel häufiger im Vergleich zu Festnetzen um ein vielfaches höhere Bitfehlerraten, die auch von der Sicherungsschicht nicht mehr ausgeglichen werden können und der Übergang des mobilen Gerätes zwischen Zugangspunkten. Ein weicher Übergang, bei dem keine Daten verloren gehen, ist kaum möglich. Bis zur Bekanntgabe des neuen Anschlußpunktes werden noch Daten an den alten gesendet, obwohl dort die Verbindung zum mobilen Gerät schon unterbrochen ist. Diese durch Mobilität begründeten Verluste können von einem klassischen TCP allerdings nicht erkannt werden, da hier ein Fehlerkontrollmechanismus, nämlich das Ausbleiben einer Bestätigung durch Übertragungsfehler zur Staukontrolle zweckentfremdet wird. Der daraus resultierende Einsatz von Slow Start bei Übertragungsfehlern oder dem Übergang zwischen Zugangspunkten ist nicht sinnvoll.

Voraussetzungen für Erweiterungen und Optimierungen von TCP für drahtlose Netze sind Kompatibilität zum klassischen TCP, da die Zahl der Endgeräte sehr groß ist und eine Beibehaltung des Stauverhaltens für den drahtgebundenen Teil der Verbindung, um die Funktionsfähigkeit des Internet nicht zu gefähren.

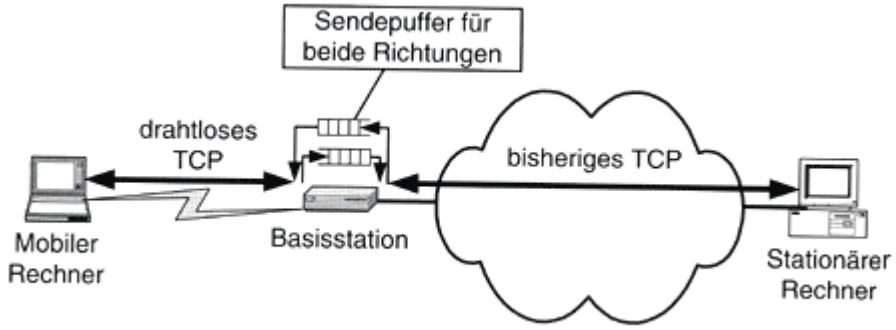
2. TCP Optimierungen

2.1. Split Connection Verfahren



Hierbei wird die Verbindung zwischen stationärem und mobilem Gerät in Festnetzteil und drahtlose Strecke unterteilt. Das Verbindungsglied beider Teilstrecken stellt eine Basisstation dar.
Für die Kommunikation zwischen stationärem Gerät und Basisstation wird hierbei klassisches TCP eingesetzt. Dadurch ist eine Anpassung des TCP auf der Seite des stationären Gerätes nicht notwendig und das Stauverhalten von TCP wird nicht beeinflußt. Auf der drahtlosen Teilstrecke kann ein optimiertes TCP eingesetzt. Aber selbst ein klassiches TCP profitiert von den geringeren Paketlaufzeiten auf dem mobilen Teilabschnitt, was eine schnellere Übertragungswiederholung im Fehlerfall möglich macht.
Durch die Zwischenschaltung der Basisstation ist der drahtlose Teil der Verbindung für den Computer im Festnetz nicht sichtbar. Als Ort der Trennung können neben einer Basisstation auch andere Netzzugangspunkte, wie spezielle Server zum Übergang in Mobilfunknetze genutzt werden.

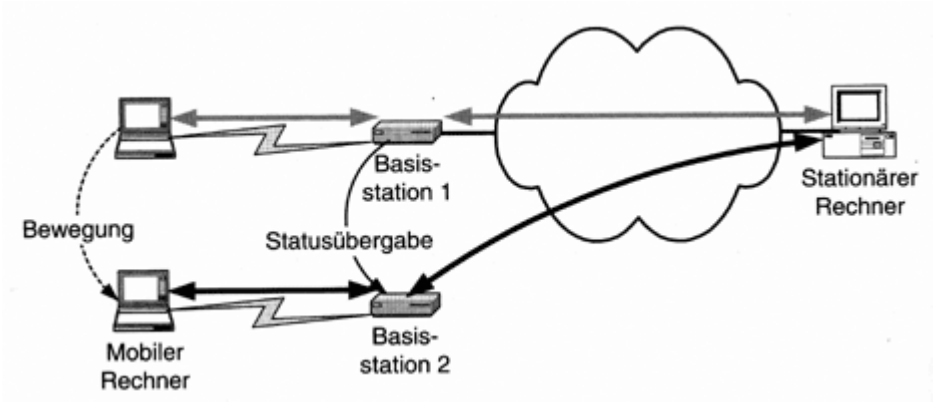
2.1.1. Indirektes TCP (I-TCP)



Bei I-TCP handelt die Basisstation als Stellvertreter bzw. Proxy und leitet Daten in beide Richtungen weiter. Der Proxy stellt außerdem das Ende der TCP-Verbindung der jeweiligen Teilstrecke dar. Dabei werden die Daten zwischengespeichert und unmittelbar Bestätigungen an den Sender geschickt. Erst dann werden die Daten zum Eigentlichen Empfänger weitergeleitet.

Übertragungsfehler auf der drahtlosen Strecke werden direkt vom Stellvertreter behandelt, indem lokal die Übertragung aus dem Zwischenspeicher wiederholt wird. Der Sender im Festnetz bleibt so von Übertragungsfehlern auf der mobilen Strecke unbeeinflußt, denn bevor dort Alarm wegen fehlender Bestätigungen ausgelöst werden kann, hat der Proxy die Übertragung bereits wiederholt. Das schnellere Feststellen eines Paketverlustes auf dem mobilen Teilabschnitt wird hier durch die geringe Umlaufdauer der Pakete möglich.

Beim Wechsel des Zugangspunktes regelt Mobile IP die Umleitung der Datenpakete zum neuen Zugangspunkt. Allerdings müssen bereits im alten Proxy zwischengespeicherte Daten und der Verbindungsstatus an den neuen Agenten/Proxy weitergeleitet werden. Der Verbindungsstatus beinhaltet Sequenznummer des zuletzt empfangenen Paketes, die IP-Adressen und die Portnummern der Anwendung.



Für den stationären Computer bleibt dieser Wechsel verborgen und die Verbindung bleibt erhalten.

Vorteile von I-TCP:

Änderungen auf der Festnetz-Seite sind nicht notwendig und auf dem drahtlosen Teilabschnitt nicht zwingend. Optimierungen, wie die Abschaltung von Slow Start wirken sich nur auf den drahtlosen Teilbereich aus. Weiterhin können sich Übertragungsfehler auf der drahtlosen Teilstrecke nicht ins Festnetz fortpflanzen, da den Stellvertreter nur reihenfolgerichtige Pakete ohne Lücken verlassen.

Und durch die geringeren Verzögerungen auf der drahtlosen Teilstrecke schneller durchführbaren Übertragungswiederholungen kann sich auch ein eingesetztes klassisches TCP schneller von Paketverlusten erholen. Wenn der Stellvertreter als Gateway agiert können auch von TCP gänzlich verschiedene Transportschicht-Protokolle eingesetzt werden. Denkbar wären bespielsweise Protokolle mit kleineren Paketköpfen, um Bandbreite zu sparen.

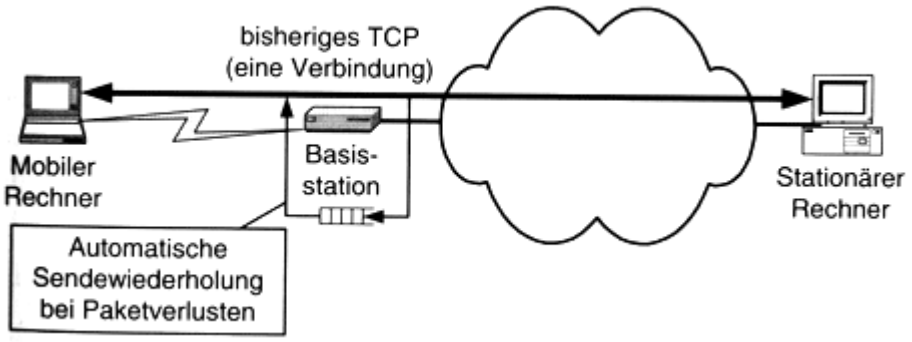
Nachteile von I-TCP:

Das Auftrennen der Verbindung hat einen Verlust der Ende-zu-Ende Semantik von TCP zur Folge. So garantiert dem Sender eine empfangene Bestätigung noch nicht den Empfang beim eigentlichen Empfänger. Durch Fehlfunktionen beim Proxy können so bereits bestätigte Pakete verloren gehen und Fehlfunktionen beim Kommunikationspartner hervorrufen.

Weiterhin tritt beim Wechsel des Zugangspunktes durch die Weiterleitung bereits zwischengespeicherter Pakete vom alten Proxy zum neuen eine starke Verzögerung auf. Während beim neuen Stellvertreter schon weitere Pakete vom Sender eintreffen muß mit der Weiterleitung zum mobilen Gerät gewartet werden, bis alle fahlenden Pakete vom alten Stellvertreter eintreffen, da sonst die Paketreihenfolge vertauscht wird und der Empfänger von Verlust ausgehen würde. Dieses Verfahren bringt auch den großen Nachteil mit sich, daß auch der neue Zugangspunkt I-TCP unterstützen muß.

Bei Einsatz von Ende-zu-Ende Verschlüsselungen muß der Stellvertreter außerdem voll einbezogen werden, da er sonst die Paketnummern nicht lesen und die richtige Reihenfolge herstellen könnte.

Snooping TCP



Snooping TCP ist eine transparente Erweiterung, da hier Pakete in Richtung des mobilen Gerätes zwar zwischengespeichert, aber nicht von der Basisstation bestätigt werden. Durch die Nähe zum mobilen Gerät ist eine schnelle Übertragungswiederholung bei Paketverlust auf dem mobilen Teilabschnitt möglich.

Auch Pakete in Richtung des stationären Computers werden mitgehört. Beim Mithören einer Empfangsbestätigung werden die bestätigten Pakete aus dem Zwischenspeicher der Basisstation gelöscht. Bei ausbleibenden oder mehrfachen Bestätigungen ist es möglich die fehlenden Pakete aus dem lokalen Zwischenspeicher der Basisstation schnell erneut zu senden, bevor der Kommunikationspartner im Festnetz das fehlen der Pakete bemerkt und selbst sie erneut sendet. Weiterhin werden doppelte Bestätigungen vor dem Weiterleiten an den stationären Computer herausgelöscht, um eine unnötige Übertragungswiederholung zu verhindern und aus dem Festnetz eintreffende Duplikate von lokal neu übertragenen Paketen verworfen.

Treten beim Senden vom mobilen Gerät ins Festnetz Lücken im Datenstrom durch Paketverluste auf dem mobilen Teilabschnitt auf, kann die Basisstation sofort das mobile Gerät darüber benachrichtigen und so eine erneute Übertragung einleiten, bevor der stationäre Computer ein Fehlen bemerkt.

Vorteile von Snooping TCP:

Die Ende-zu-Ende Semantik von TCP bleibt erhalten. So spielen Fehlfunktionen beim Snooping TCP Agenten keine Rolle und bleiben unbemerkt, weil in diesem Falle traditionelles TCP weiter eingesetzt wird. So ist auch bei beiden Kommunikationsteilnehmern keine Änderung am TCP notwendig.

Beim Wechsel des Zugangspunktes ist keine Zustandsübergabe und Weiterleitung das Zwischenspeichers notwendig. Im ungünstigsten Fall müssen noch an den alten Agenten geleitete Pakete vom Sender erneut übertragen werden. Dadurch ist es auch nicht notwendig, daß der neue Agent Snooping TCP unterstützt.

Nachteile von Snooping TCP:

Durch die geringe Isolation der mobilen Verbindung wird bei größeren Störungen oder Überlastungen auch die Kommunikation auf dem Festnetz-Teil in Mitleidenschaft gezogen und stark gebremst.

Wenn negative Bestätigungen auf dem drahtlosen Teil eingesetzt werden sollen ist dafür eine Anpassung des TCP auf dem mobilen Gerät nötig.

Bei Einsatz von Ende-zu-Ende Verschlüsselungen schlägt das mitlesen der Pakete fehl und die Verbindung fällt auf Standard TCP zurück. Viele zeitabhängige Sicherheitsmechanismen verhindern außerdem das Wiederholen von gleichen Datenpaketen.
Der Einsatz von Verschlüsselungsmechanismen auf Anwendungsschicht (SSL/TLS) ist allerdings ohne Probleme möglich, da hier nicht die Header der Pakete mitverschlüsselt werden.

Mobile TCP (M-TCP)

Dient der Optimierung des TCP Verhaltens bei längeren Verbindungsunterbrechnungen beispielsweise bei Lücken in der Netzabdeckung. Bei I-TCP müßen hier immer mehr Daten zwischengespeichert und ggf. an einen neuen Proxy weitergeleitet werden. Beim Einsatz von traditionellem TCP würde der Sender stetig versuchen die nicht bestätigten Pakete erneut zu übertragen, dabei Slow Start einsetzen und schließlich die Verbindung ganz abbrechen.

Bei Mobile TCP wird die Verbindung ähnlich wie bei I-TCP aufgeteilt. Anstelle des Proxies tritt ein Überwachungsknoten (Supervisory Host), welcher alle Pakete zum mobilen Gerät und alle Bestätigungen beobachtet aber nicht die Daten zwischenspeichert. Gehen Pakete verloren müssen sie vom Sender wiederholt werden.

Auf der Teilstrecke zum mobilen Gerät kommt ein optimiertes TCP zum Einsatz. Bleiben Bestätigungen vom mobilen Gerät aus, wird davon ausgegangen, daß gerade die Verbindung unterbrochen ist. Nun drosselt der Überwachungsknoten den stationären Rechner, indem sein Sendefenster auf 0 gesetzt wird (Persistenter Modus). So wird der Sender seinen aktuellen Zustand beibehalten und nicht versuchen erneut Daten zu übertragen, bis wieder eine Verbindung zum mobilen Gerät besteht und der Überwachungsknoten die Verbindung eröffnet, indem das Sendefenster auf den alten wert gesetzt wird.

Vorteile von Mobile TCP:

Auch hier bleibt die Ende-zu-Ende Semantik von TCP erhalten, da der Überwachungsknoten selbst keine Bestätigungen versendet.

Auch sind keine Änderungen beim TCP des stationären Rechner nötig bei gleichzeitigem Einsatz von optimiertem TCP ohne Slow Start auf dem drahtlosen Teil. Ein Verzicht auf Slow Start macht allerdings den Einsatz eines Bandbreitenverwalters nötig.

Unnötige Übertragungswiederholungen werden vermieden, wenn der mobile Rechner getrennt ist. Eine Zustandsübergabe beim Wechsel des Zugangspunktes ist nicht nötig, da keine Zwischenspeicherung erfolgt.

Nachteile von M-TCP:

Da geringe Übertragungsfehler auf dem drahtlosen Teil angenommen werden, sind durch die fehlende Zwischenspeicherung Paketverluste auf der drahtlosen Strecke auch für den stationären Rechner sichtbar.

Auf allen mobilen Teilnehmern ist der Einsatz des modifizierten TCP notwendig, was eine Zugangsbarriere darstellt.
Das Fehlen von SlowStart macht einen Bandbreitenverwalter notwendig.

Fast Retransmit / Fast Recovery
beim Wechsel zu neuen Zugangspunkt

Dieses Verhalten kann künstlich nach dem Wechsel des Zugangspunktes durch das Versenden duplizierter Bestätigungen seitens des mobilen Gerätes hervorgerufen werden. Dies zwingt den Kommunikationspartner sofort Fast Retransmit einzusetzen und auf Slow Start zu verzichten. Somit wird die Übertragung mit der vorherigen Senderate fortgesetzt.

Vorteile:

Es sind nur relativ kleine Software-Änderungen am mobilen Gerät notwendig. Weder Zugangspunkt noch Kommunikationspartner im Festnetz müssen modifiziert werden.

Nachteil:

Die Einfachheit bringt wieder die Nachteile der ungenügenden Isolation des mobilen Teilabschnittes bei Paketverlusten mit sich. Paketverluste durch Übertragungsfehler werden nicht berücksichtigt. Übertragungswiederholungen müssen vom Sender im Festnetz initiiert werden.
Bei länger andauernden Wechseln verfällt der Sender in Slow Start und beginnt mit dem erneuten Senden der Daten.

Einfrieren der Verbindungsparameter

Bei längeren Unterbrechnungen kann die MAC Schicht TCP über bevorstehende Unterbrechnungen informieren, damit nicht fälschlicherweise von einem Stau ausgegangen wird. TCP wird daraufhin das Senden einstellen und den bisherigen Zustand beibehalten. Wenn die Verbindung noch lang genug besteht, kann auch der andere Kommunikationsteilnehmer über die Art der Unterbrechnung benachichtigt werden.

Ist die Verbindung wiederhergestellt, informiert die MAC Schicht TCP und die Übertragung wird mit den alten Parametern fortgesetzt.

Verbindungen können auch nach langen Unterbrechnungen fortgesetzt werden. Der Einsatz ist unabhängig von anderen TCP Mechanismen, wie z.B. Verschlüsselungen. Allerdings muß die Software auf beiden Kommunikationsteilnehmern angepaßt werden.

Selektive Übertragungswiederholung

Hierbei kann die Wiederholung einzelner Pakete indirekt angefordert werden, da eine Möglichkeit zur Bestätigung einzelner Pakete geschaffen wird. Nur die fehlenden Pakete werden erneut übertragen.

Dies wird bereits von vielen TCP Implementierungen unterstützt. Allerdings ist die Software auf Empfängerseite etwas komplexer, da Pakete zwischengespeichert und selbst in die richtige Reihenfolge gebracht werden müssen.