Echtzeit–Datenverkehr über IP–basierte Datennetze

Ursula Hilgers, Falko Dressler

Regionales Rechenzentrum der Universität Erlangen–Nürnberg, Martensstraße 1, 91058 Erlangen

Zusammenfassung Über die IP–Infrastruktur werden heute schon Applikationen mit unterschiedlichsten Anforderungen an die Kommunikationsinfrastruktur übertragen. Diese Arbeit gibt einen Überblick über die Mechanismen, die in IP–Netzwerken zur Bereitstellung von Dienstgüte zur Verfügung stehen bzw. sich in Standardisierungsgremien in Entwicklung befinden. Darüber hinaus sollen die Möglichkeiten moderner IP–Datennetze in Hinblick auf zeitkritisches Echtzeit–Verhalten betrachtet werden. Unter diesen Gesichtspunkten werden Netzwerkkomponenten in einfachen Testbeds untersucht. Die Ergebnisse sollen einen Überblick über die aktuell verfügbaren Möglichkeiten geben, Echtzeit–Daten über das IP–Protokoll zu übertragen.

1 Einleitung

Gerade für die Anforderungen von Echtzeit–Systemen in der Automatisierungstechnik gibt es schon seit vielen Jahren proprietäre Lösungen, lokale Vernetzungen unter Einhaltung von Echtzeit–Kriterien aufzubauen. Ein wesentliches Ziel der heutigen Entwicklung ist es, eine Migration aller Dienste auf ein einziges zugrunde liegendes Datennetz zu erreichen. So sollen und werden auch schon heute IP–basierte Netzwerke für den Datenaustausch von Echtzeit–Geräten genutzt. Dies aber nur für wenig zeitkritische Offline-Datentransfers.

Dieser Beitrag verfolgt zwei Ziele. Zum einen soll ein Überblick über Mechanismen gegeben werden, die in IP–Netzwerken zur Bereitstellung von Dienstgüte zur Verfügung stehen. Auf der anderen Seite sollen die Möglichkeiten moderner IP–Datennetze in Hinblick auf zeitkritisches Echtzeit–Verhalten untersucht werden. Die Mechanismen werden in einfachen Testbeds auf ihre Fähigkeit untersucht, Applikationen mit Echtzeit–Anforderungen zu übertragen. Für die Messungen stehen Router der Firma Cisco zur Verfügung.

2 Echtzeit–Datenübertragung

Bei „klassischen“ Echtzeitanforderungen von Applikationen muss ein Datum innerhalb einer vorgeschriebenen Zeitspanne beim Empfänger ankommen, da es sonst wertlos ist bzw. eine wichtige Aktion nicht zeitgerecht durchgeführt werden kann [SN95]. Wird die Übertragung von Daten über IP–Netzwerke untersucht, ist nicht nur die maximale Verzögerung interessant, sondern vielmehr der durch das IP–Netzwerk erzeugte Jitter (Variation des Delays). Verlangt wird eine vorhersagbare, schnelle Antwortzeit für zeitkritische Ereignisse mit einer exakten Zeitinformation. Die Bearbeitung solcher Ereignisse muss streng prior durchgeführt werden und die Netzwerkkomponenten müssen sich stabil unter extremer Last verhalten. Die vorgegebene Bandbreite muss für die Dauer der Bearbeitung garantiert werden, und die Fehlerrate sollte so klein wie möglich sein. Zusätzlich ist eine 100 % Verfügbarkeit unbedingt notwendig.

Die *Internet Engineering Task Force* (IETF) unterscheidet zwischen zwei grundsätzlich verschiedenen Typen von Applikationen. Zum einen soll ein *Controlled–Load Service* [Wro97] für Anwendungen mit weniger harten Echtzeitanforderungen vorhanden sein, welche typischerweise im Multimedia–Bereich zu finden sind. Ein gewisses Maß an Paketverlusten und Delay wird durch adaptive Multimedia–Algorithmen, die auf zu geringe Qualität einer Übertragung z.B. durch Reduktion der Bandbreite reagieren können, akzeptiert. So leidet die subjektive Qualität bei der Durchführung einer Videokonferenz nicht unbedingt durch den Verlust eines einzelnen Videoframes. Leichtes Verletzen der zeitlich vorgegebenen Schranken hat keine katastrophalen Folgen, wobei diese Aussage sicherlich nicht generell zutrifft. Der zweite Typ ist der sogenannte *Guaranteed Service* [She97], welcher für Applikationen gedacht ist, die weniger fehlertolerant sind. Ein Beispiel für harte Echtzeit–Anforderungen mit hoher Verfügbarkeit sind hochqualitative Videoübertragungen in der Telemedizin.

Werden diese Anforderungen auf die Realisierung in Netzwerkkomponenten übertragen, bedeutet es, dass zunächst Pakete als zu einer Echtzeit–Applikation gehörend gekennzeichnet werden müssen. Dann muss für diese Pakete Bandbreite in den Netzwerkkomponenten und auf den physikalischen Verbindungen reserviert werden. Zeitliche Anforderungen müssen eingehalten werden. Um die Verfügbarkeit zu steigern, sollten alternative Wege im Netzwerk zur Verfügung stehen. Die dadurch höhere Verfügbarkeit erkauft man sich aber durch eine (zeitweise) Vergrößerung von Delay, Jitter und Paketverlustrate.

3 Echtzeit–Datenübertragung in IP–Netzwerken

Um in IP–Netzwerken verschiedene Dienstklassen unterstützen zu können, muss die Standard–Funktionalität der Netzwerkkomponenten erweitert werden. Neben der Klassifikation von Verkehr sind Algorithmen zum Ressourcen–Management, zur Verhinderung von Überlast und Scheduling–Mechanismen zur Bereitstellung von unterschiedlichen Dienstklassen notwendig. Eine Bewertung der Mechanismen zur Verkehrsüberwachung und –regulierung ist in [Hil00] zu finden.

3.1 Klassifikation

Um Echtzeit–Verhalten im Netzwerk bereitzustellen, müssen IP–Pakete in unterschiedliche Klassen eingeteilt werden können. Das IP–Protokoll verfügt zunächst über keinerlei Möglichkeit der Verkehrsdifferenzierung. Es existiert nur eine Best Effort Dienstklasse. Allerdings kann das *Diensttypfeld* im IP–Header, das *Type of Service* (ToS) Feld, dazu verwendet werden, IP–Pakete verschiedenen Dienstklassen zuzuordnen. [DA81a] unterteilt das ToS–Feld in drei *Precedence*–Bits, drei ToS–Bits und zwei weitere Bits, die nicht verwendet werden.

Die Klassifikation des IP–Verkehrs erfolgt meist auf der Basis von *Anwendungsflüssen* (*Flows*). Dabei gehören alle Pakete mit den gleichen Sende– und Empfangsadressen sowie den gleichen Sende– und Empfangsportnummern zu dem gleichen Flow [SW97].

3.2 Algorithmen zur Verhinderung von Überlast

Random Early Detection (RED) [FJ93], [Hil00] ist ein Mechanismus zur Verhinderung von Überlast in Netzwerken. Durch das kontinuierliche Beobachten der durchschnittlichen Länge der Warteschlange an jedem Ausgangs–Interface der Vermittlungsrechner kann der Beginn einer Überlastsituation daran erkannt werden, dass der durchschnittliche Füllgrad der Warteschlange ansteigt. Der Router kann reagieren, bevor die Warteschlange überläuft, indem er den Füllgrad der Queue reduziert. Dabei werden zufällig ausgewählte Pakete verworfen, unabhängig von der Verbindung, zu der sie gehören, sodass die Wahrscheinlichkeit, dass Pakete einer Verbindung gelöscht werden, proportional ist zu dem Anteil an Ausgangsbandbreite an dem Router–Interface, den diese Verbindung für sich beansprucht. Das bedeutet, dass durch das Verfahren RED Verbindungen benachteiligt werden, die durch ihr hohes Verkehrsaufkommen andere Verbindungen mit niedrigen Bandbreitenanforderungen beeinträchtigen.

Protokolle wie z.B. das TCP/IP Protokoll reagieren auf verworfene Pakete durch das Reduzieren ihre Übertragungsbandbreite. Durch das dedizierte Verwerfen von Paketen einzelner TCP–Verbindungen kann der Effekt der globalen Synchronisation, bei der alle TCP–Verbindungen ihre Übertragungsrate reduzieren, verhindert werden, wobei der Gesamtdurchsatz im Netzwerk ansteigt. Außerdem wird bei der Aktivierung von RED das Delay von Datenpaketen durch den geringen Füllgrad der Warteschlangen klein gehalten.

3.3 Scheduling

Scheduling–Mechanismen werden an ausgehenden Interfaces von Routern aktiviert. Sie bestimmen die Bandbreite, die den Paketen einer Verbindung zugewiesen wird, die Reihenfolge, mit der sie bedient werden, und die Menge des Speicherbereichs, der ihnen zur Verfügung steht. Damit haben sie einen Einfluss darauf, wie unterschiedliches Verhalten in den Netzwerkkomponenten zur Realisierung von verschiedenen Dienstklassen implementiert werden kann.

Das in Routern per Standard aktivierte Scheduling–Verfahren FiFo Queueing (First in First out) ist nicht geeignet, Pakete mit sich unterscheidender Priorität weiterzuleiten oder ihnen einen unterschiedlichen Anteil an Bandbreite an einem ausgehenden Router–Interface zuzuteilen. Ein Scheduling–Mechanismus zur Implementation verschiedener Dienstklassen ist das Weighted Fair Queueing (WFQ). Der Scheduling–Mechanismus verwaltet mehrere Warteschlangen für unterschiedliche Dienstqualitäten. Jeder Warteschlange wird ein Gewicht zugeordnet, das den Anteil an ausgehender Bandbreite festlegt. Damit kann für jede Dienstklasse, der eine ausgehende Warteschlange zugeteilt wird, eine untere Grenze für den Durchsatz angegeben werden, die nicht unterschritten wird. Wird der Verkehr einer Klasse zusätzlich durch einen Token Bucket reguliert, kann eine maximale obere Grenze für das Ende–zu–Ende–Delay angegeben werden [Par92].

3.4 Ressourcen–Management

Zur Realisierung von hochprioren Dienstklassen müssen die Ressourcen in einem Netzwerk verteilt und verwaltet werden. Dazu wird bei verbindungsorientierten Protokollen zunächst bei Verbindungsaufbau durch eine Call Admission Control (CAC) überprüft, ob ausreichende Ressourcen zu Verfügung stehen, um der Verbindung die geforderte Dienstgüte auf dem gesamten Weg durch das Netzwerk während ihrer Lebenszeit

zu garantieren. Dann müssen die Ressourcen in den Netzwerkkomponenten reserviert werden. Aufsetzend auf dem IP–Protokoll ist — zur Zeit — noch kein Mechanismus zur Ressourcen–Reservierung implementiert, der auf Flows–Basis angewendet werden kann. Ein Ansatz der IETF ist das *Resource Reservation Setup Protocol* (RSVP) [Bra97]. Es hat sich allerdings herausgestellt, dass dieses Protokoll in WANs nicht skaliert [Man97], da es für jeden Flow Reservierungszustände in den Knoten und den Endgeräten verwalten muss.

4 Empirische Messungen in IP–Netzwerken

Im Folgenden sollen einige der bereits im vorhergehenden Abschnitt erwähnten Funktionalitäten zur Paket–Klassifikation, zur Verhinderung von Überlast und zum Scheduling an Messungen in einfachen Testbeds untersucht werden.

4.1 Klassifikation

Um Echtzeit–IP–Pakete mit einer bestimmten Eigenschaft, z. B. der gleichen Zieladresse, in eine gemeinsame Dienstklasse einzuordnen, müssen alle Pakete zunächst nach dieser Eigenschaft gefiltert werden. Diese Operation wird bei Routern der Firma Cisco mit sogenannten *Access Control Listen* (`ACL`) durchgeführt. Sie kontrollieren an einem Interface eines Routers alle Pakete, die empfangen werden. Pakete mit gleichen Eigenschaften werden herausgefiltert, um sie durch Setzen der Precedence–Bits im IP–Header einer Dienstklasse zuzuweisen. Das Markieren, d. h. das Setzen der Bits im IP–Header, wird von Cisco Routern mit dem Mechanismus *Committed Access Rate* (`CAR`) realisiert.

Die Auswertung der Filterregeln kann auf Routern in Software oder in Hardware implementiert sein, was einen entscheidenen Einfluss auf die Performance dieser Operation hat. Bei den folgenden Untersuchungen wird die Klassifikation von IP–Paketen anhand der Zieladresse durchgeführt. Die Pakete werden in Abhängigkeit der Adresse in fünf verschiedene Klassen eingeordnet, wobei die Zuordnung zu einer Dienstklasse durch entsprechendes Markieren der Precedence–Bits im IP–Header erfolgt. Im Laufe der Messung werden die IP–Pakete mit unterschiedlicher Zieladdresse und somit auch die `ACLs` gesteigert, sodass ihr Einfluss auf die Funktionsweise des Router–Interfaces untersucht werden kann. Mit dieser Messung soll gezeigt werden, dass mit einer zunehmenden Zahl von ACLs die Prozessorlast am eingehenden Interface des Routers so sehr

steigt, dass er seine Hauptaufgabe, das Weiterleiten von Paketen, nicht mehr durchführen kann [HH01].

Der Testaufbau ist in Abbildung 1 a) abgebildet. Es wird ein Cisco 7507 Router mit Betriebssystem 12.1(3a)T1 verwendet. Der Router ist mit Versatile Interface Prozessor (VIP) Boards bestückt, die die Paketfilterung und das Klassifizieren selbständig durchführen, ohne dass Teile der Aufgabe vom Zentralprozessor ausgeführt werden. Zwei dieser Boards sind mit einem Verkehrsgenerator und –analysator, einem Smartbits 6000, über zwei STM1 Packet over SONET (PoS) Verbindungen verbunden [Gor97]. Über das eine Interface empfängt der Router die Pakete vom Verkehrsgenerator, über das andere sendet er die Pakete wieder zum Analysator zurück. Die Funktionalität des Klassifizierens ist am eingehenden Interface des Routers aktiviert. Der Verkehrsgenerator sendet über eine kontinuierlich erhöhte Anzahl von Flows Pakete mit unterschiedlichen Ziel–IP–Adressen. Die IP–Pakete haben eine Größe von 429, der durchschnittlichen Paketgröße im Deutschen Wissenschaftsnetz.

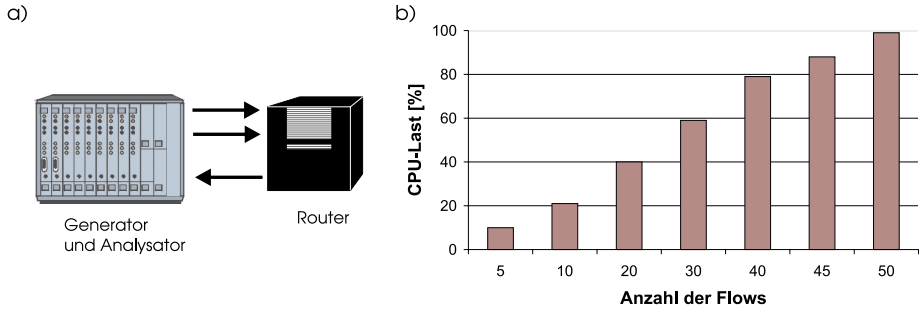

Abbildung1. a) Messaufbau zum Test, welche Auswirkungen die Klassifikation auf die CPU–Belastung am eingehenden Interface eines Routers hat, b) Abhängigkeit der CPU–Belastung des eingehenden VIP Boards von der Anzahl der klassifizierten Flows.

Abbildung 1 b) zeigt das Ergebnis des Versuchs. Auf der Abszisse ist die Anzahl der Flows dargestellt, die während der Messung von 5 auf 50 gesteigert wird [Hil01]. Auf der Ordinate ist die CPU–Last des eingehenden Router–Interfaces in Prozent aufgetragen. Auf den Flows wird mit einer Rate von 2,9 MBit/s gesendet. Empfängt der Router an seinem eingehenden Interface fünf Flows, deren IP–Pakete markiert werden, wird eine CPU–Belastung auf dem eingehenden Interface des Routers von 15 % beobachtet. Die Last steigt bis zu 88 % bei 45 Flows und bei 50

Flows gehen bei einer CPU–Belastung von 100 % sogar Pakete verloren. Als Vergleich sei die CPU–Last genannt, wenn mit dem beschriebenen Testaufbau und dem beschriebenen Sendeverhalten die ACLs nicht konfiguriert sind: Bei 10 Flows steigt die CPU auf 10 %, bei 20 Flows auf 21 %. Da in realen Netzen mehrere tausend Flows an einem Router–Interface in einem Netzwerk auflaufen, zeigen diese Ergebnisse kein wünschenswertes Verhalten.

4.2 Algorithmen zur Verhinderung von Überlast

Die im Abschnitt 3.2 erläuterte Eigenschaft, dass bei der Aktivierung von RED das Delay reduziert wird durch den geringeren Füllgrad der Warteschlangen, soll im Folgenden anhand von Labormessungen untersucht werden. Verglichen wird das Delay bei der Datenübertragung mit RED mit den entsprechenden Tests ohne Aktivierung von RED.

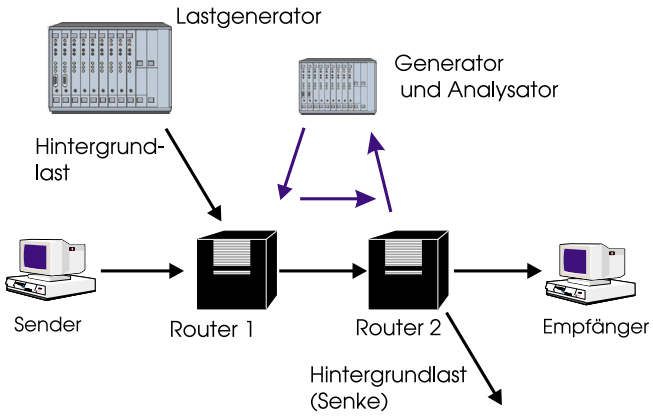

Abbildung2. Messaufbau zur Untersuchung der Funktionweise von RED.

Abbildung 2 zeigt den Testaufbau. Zwei SUN Ultra 1 Solaris–Workstations, `Sender` und `Empfänger`, sind jeweils mit FDDI mit zwei Cisco 7507 Routern verbunden. Die beiden Router mit IOS 11.1-24.CC verbindet eine STM1 PoS Verbindung. Ein ATM–Monitor HP E4200B ist mit `Router1` verbunden und sendet UDP–Verkehr als Hintergrundlast über zwei STM1 Verbindungen. Die Last wird von `Router2` abgeleitet. Die beiden Workstations tauschen TCP–Pakete untereinander aus. Zusätzlich sendet ein weiterer Generator mit Zeitstempel versehene UDP–Pakete mit

einer Rate von 0,1 MBit/s zum `Router1`, die über `Router2` an den Verkehrsgenerator zurückgesendet werden. Aus den Zeitstempeln in diesen Paketen kann die Verzögerung in den Routern bestimmt werden.

Abbildung 3 zeigt das durchschnittliche Delay in Abhängigkeit von der Hintergrundlast. Deutlich ist zu sehen, dass das Delay bei höherer Hintergrundlast kleiner ist als ohne RED. Außerdem wächst es ab einer Last von 75 MBit/s nicht weiter an. Dies ist durch die geringere durchschnittliche Länge der Warteschlange am Ausgangs–Interface des Routers begründet. Den Vorteil des geringeren Delays erkauft man sich allerdings durch eine höhere Paketverlustrate, bei mittlerer Auslastung des Interfaces.

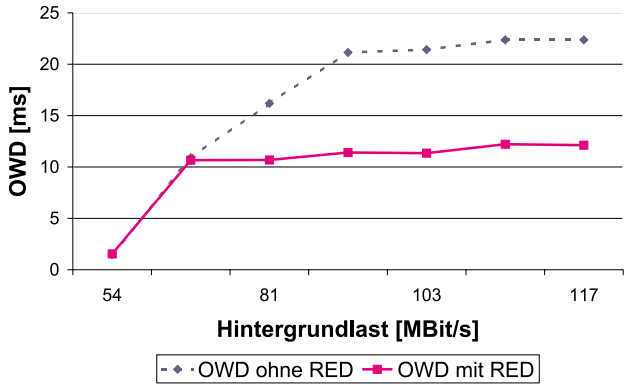

Abbildung3. Messung des Delays mit und ohne Aktivierung von RED bei steigender Hintergrundlast.

4.3 Scheduling

Um hochprioren Verkehr, z. B. Echtzeit–Daten, entsprechend der zeitlichen Anforderungen weiterzuleiten, existiert auf den Cisco Routern GSR12008 ein Scheduling–Verfahren, das die strikte Priorisierung von Paketen einer Warteschlange unterstützt (*Low Latency Queueing*). Der Testaufbau, mit dem dieses Verfahren untersucht werden soll, ist in Abbildung 4 a) dargestellt. Die beiden Router mit Software–Version 12.0(9)S sind über ein STM4 Interface verbunden und zusätzlich mit je zwei STM4 und einem STM1 Interface mit einem Verkehrsgenerator bzw. –analysator gekoppelt. Um ein Netzwerk mit drei Dienstklassen zu simulieren, werden

auf den beiden STM4 Verbindungen mit einer Rate von 593 MBit/s 429 Byte große Pakete gesendet, auf der STM1 Verbindung Pakete von 240 Byte mit einer Rate von 113,25 KBit/s, die einem hochprioren Sprachdatenstrom entsprechen könnten. Alle Ströme werden über die gleiche Verbindung von `Router1` zu `Router2` geleitet, sodass dort eine Überlastsituation entsteht. Die Konfiguration der Warteschlangen für die drei Ströme am ausgehenden Interface von `Router1` ist in Abbildung 4 b) dargestellt: Der hochpriore Datenstrom, der nur eine geringe Bandbreite erfordert, wird durch die streng priorisierte Warteschlange mit einem Anteil an Ausgangsbandbreite von 5 % weitergeleitet, die beiden niederprioren Warteschlangen mit je einen Anteil von 40 %. Auf ihnen ist zusätzlich noch das Verfahren RED konfiguriert. Mit Hilfe des Verkehrsanalysators sollen der Durchsatz, das Delay und der Jitter der Pakete in der hochprioren Warteschlange bestimmt werden, wobei die Ergebnisse beim FiFo Queueing mit denen bei Aktivierung von Low Latency Queueing verglichen werden.

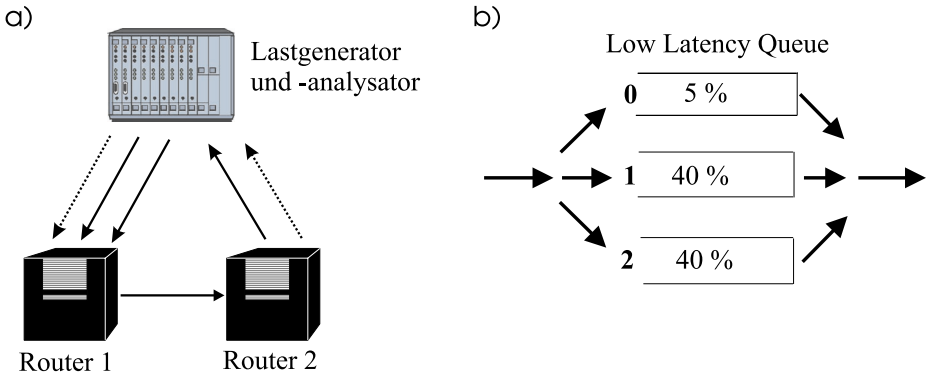

Abbildung4. a) Messaufbau zum Vergleich von FiFo Queueing und Priority Queueing, b) Konfiguration bei der Messung mit Priority Queueing.

Zunächst wird der Durchsatz des hochprioren Stroms mit FiFo–Scheduling überprüft. Es zeigt sich, dass durch die Überlast am ausgehenden Interface des `Router1` nur ein Drittel aller Pakete des Stroms das Ziel mit einem Delay von 302 ms erreichen. Bei Aktivierung von Low Latency Queueing reduziert sich das Delay auf 2,66 ms und alle Pakete erreichen das Ziel. Auch der maximale Jitter reduziert sich von 30,12 ms bei FiFo Queueing auf 0,049 ms. Damit zeigt sich, dass das Verfahren zur

Weiterleitung von Echtzeit–Verkehr geeignet ist, da der hochpriore Strom bevorzugt bearbeitet und weitergeleitet wird.

5 Messungen in realen Netzwerken

Im Rahmen einer Diplomarbeit wurde am Regionalen Rechenzentrum der Universität Erlangen ein Programm entwickelt, das Leistungsmessungen in realen Netzwerken durchführen kann [Hof01]. Das Programm bestimmt das One–Way–Delay und den Jitter auf Verbindungen und wurde im Deutschen Wissenschaftsnetz im Dezember 2000 auf der Verbindung Erlangen — Berlin eingesetzt.

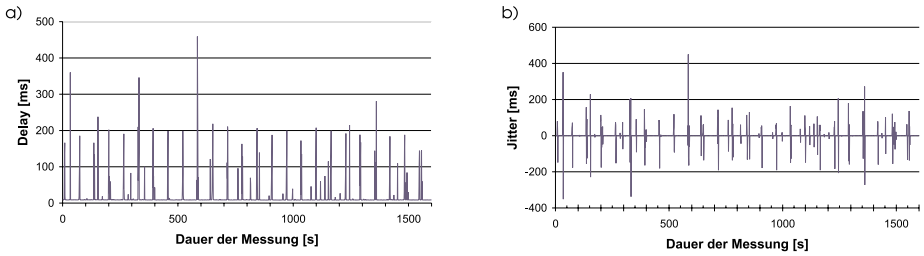

Abbildung5. Messungen von Delays und Jitter.

In Abbildung 5 a) ist das Delay auf dieser Verbindung abgebildet. Das Delay schwankt zwischen 8 und 460 ms. Abbildung 5 b) gibt Aufschluss über den Jitter. Er schwankt zwischen -350 und 450 ms. Dabei konnte durch Referenzmessungen am System selber festgestellt werden, dass der Jitter, der durch das Betriebssystem verursacht wird, vernachlässigbar ist.

Es ist offensichtlich, dass diese Werte für die Übertragung von Echtzeit–Daten nicht ausreichen.

6 Zusammenfassung

Die Ergebnisse in diesem Beitrag haben gezeigt, dass theoretische Vorarbeiten gleistet sind und auch bereits einige Mechanismen auf IP–Netzwerkkomponenten implementiert sind, die die Bereitstellung von Dienstklassen mit unterschiedlichen qualitativen Eigenschaften unterstützen. Allerdings muss festgestellt werden, dass von der Übertragung

von Echtzeit–Verkehr über IP–Netzwerke abgeraten werden muss, vor allem auch deshalb, weil noch keine Ressourcen–Management Mechanismen existieren. Interessant ist u.a. die Eigenschaft, dass durch die Aktivierung von QoS–Mechanismen ein Performance–Verlust auf den Routern hervorgerufen wird, der dazu führt, dass einzelne Datenpakete schlechter bedient werden, als der Best Effort Datenverkehr. Dafür erhalten Pakete anderer Verbindungen ein besseres Zeitverhalten.

Literatur

[Bra97] B. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, *Resource ReSerVation Protocol (RSVP) — Version 1 Functional Specification.* Request for Comments 2205, September 1997.

[DA81a] Defense Advanced Research Projects Agency, *Internet Protocol.* Request for Comments 791, September 1981.

[FJ93] S. Floyd, V. Jacobson, *Random Early Detection Gateways for Congestion Avoidance.* In: IEEE/ACM Transactions on Networking, August 1993.

[Gor97] W. J. Goralski, *SONET.* McGraw–Hill, 1997.

[HH01] U. Hilgers, R. Hofmann, *QoS — ATM versus Differentiated Services.* In: J. Knop, P. Schirmbacher (Hrsg.), Proceedings EUNIS 2001, März 2001.

[Hil00] U. Hilgers, R. Hofmann, P. Holleczek, *Differentiated Services — Konzepte und erste Erfahrungen.* In: Praxis der Informationsverarbeitung und Kommunikation, Februar 2000.

[Hil01] U. Hilgers, S. Naegele–Jackson, P. Holleczek, R. Hofmann, *Bereitstellung von Dienstgüte in IP– und ATM–Netzen als Voraussetzung für die Video–Übertragung mit Hardware Codecs.* 15. DFN–Arbeitstagung über Kommunikationsnetze, Düsseldorf, Juni 2001.

[Hof01] G. Hofmann, *Implementation eines Programms zur Bestimmung der Dienstgüte in IP–Netzen.* Diplomarbeit am Regionalen Rechenzentrum der FAU Erlangen–Nürnberg, März 2001.

[Man97] A. Mankin, F. Baker, B. Braden, S. Bradner, M. O'Dell, A. Romanow, A. Weinrib, L. Zhang, *Resource ReSerVation Protocol (RSVP) — Version 1 Applicability Statement — Some Guidelines on Deployment.* Request for Comments 2208, September 1997.

[Par92] A. K. J. Parekh, *A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks.* LIDS-TH-2089, MIT Laboratory for Information and Decision Systems, Cambridge, Mass., February 1992.

[She97] S. Shenker, C. Partridge, R. Guerin, *Specification of Guaranteed Quality of Service.* Request for Comments 2212, September 1997.

[SN95] R. Steinmetz, K. Nahrstedt, *Multimedia: computing, communications, and applications.* Prentice Hall, 1995.

[SW97] S. Shenker, J. Wroclawski, *Network Element Service Specification Template.* Request for Comments 2216, September 1997.

[Wro97] J. Wroclawski, *Specification of the Controlled–Load Network Element Service.* Request for Comments 2211, September 1997.