DFN-Projekt

Multicast-Distribution-System – MDiS

Abschlußbericht

15.2.2002

Erstellt durch:

Deutscher Wetterdienst
Kaiserleistr. 42
63067 Offenbach
Herr Knottenberg
Herr Kiehl
Herr Löser

Durchführung des Projektes in Zusammenarbeit mit:

Tellique
Kommunikationstechnik GmbH
Berliner Straße 26
D-13507 Berlin

Universität Bremen
Technologiezentrum Informatik
Bereich Digitale Medien und Netze
Bibliotheksstraße, MZH 5180
D-28359 Bremen

Das Vorhaben wurde aus Mitteln des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF) durch den Verein zur Förderung eines Deutschen Forschungsnetzes e. V. (DFN-Verein) finanziert.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

1 INHALT

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

2 EINLEITUNG

2.1 PROBLEMSTELLUNG

Durch die hohe Popularität des Internets, die steigende Bedeutung von Informationsaustausch in der Wirtschaft und die Entwicklung bandbreitenintensiver Verfahren zur Übertragung von Multimediadaten ist die effektive Nutzung der vorhandenen Netzkapazitäten ein entscheidender wirtschaftlicher und technischer Faktor.

Multicasting, die simultane Datenübertragung an mehrere, ausgewählte Empfänger, stellt daher auf Grund der resultierenden Bandbreiten- und Kostenreduzierung eine Form des Internet-Datenaustausches dar, die immer mehr an Bedeutung gewinnt und zukünftig ein zentraler Service in der Internetlandschaft sein wird.

Die notwendige Netzinfrastruktur zur Übertragung von Multicast-Paketen ist derzeit bereits verfügbar und kann z. B. zur ungesicherten Übertragung von Audio- oder Videodaten problemlos genutzt werden. Kern des Multicast-Backbones in Deutschland bildet dabei das Deutsche Wissenschaftsnetz (WiN/B-WiN).

Trotz der vorhandenen Infrastruktur und der vielfältigen Vorteile von Multicast – Verringerung der Netzlast, Reduzierung der Übertragungsverzögerung und deutlich geringere Belastung der sendenden Rechner – nutzen bisher nur wenige Dienste die Übertragung von Dateien über Multicast. Grund dafür ist, dass die gesicherte Datenübertragung oberhalb des Best-Efford-IP-Dienstes immer noch sehr aufwendig ist.

Die technische Grundlage für den Einsatz von Multicasting auf den Protokollebenen bis zur Vermittlungsschicht (ISO/OSI Ebene 3) ist heute schon vollständig gegeben. Auch ATM bietet mit Punkt-zu-Mehrpunkt-VCs die notwendigen Voraussetzungen, um beispielsweise unterhalb der LAN-Emulation 1.0 grundlegende Multicast-Funktionalität zu realisieren. Der Einsatz von Multicast wird jedoch dadurch eingeschränkt, dass eine standardisierte, universell verfügbare und problemlos benutzbare Transportplattform für die gesicherte Übertragung von Daten über Multicast (Reliable Multicast) fehlt. Marktfähige Software für die gesicherte Übertragung von Dateien über Internet liegt nur begrenzt vor und weist oftmals erhebliche Nachteile gegenüber den Unicast-Protokollen auf.

Die in den USA gegründete IP Multicast Initiative, der auch Großunternehmen wie Microsoft, Netscape und Cisco angehören, konzentriert ihre Aktivitäten auf Gruppenkommunikation und die ungesicherte Übertragung von Audio- und Videodaten. Innerhalb der IETF, dem Standardisierungs-„Gremium“ des Internets, werden zugleich in mehreren Arbeitsgruppen und in einer speziellen Researchgruppe (IRTF) gezielt Verfahren zur gesicherten Datenübertragung diskutiert. Ergebnisse dieser Gruppe oder gar entsprechende Internet-Standards können jedoch erst mittelfristig erwartet werden.

Bisher steht Anwendungen, die von den Vorteilen einer Multicast-Übertragung profitieren wollen, als standardisiertes Transportprotokoll der

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Schicht 4 nur UDP mit der von ihm angebotenen ungesicherten Übertragung von Datagrammen zur Verfügung. Viele Anwendungen sind allerdings darauf angewiesen, dass die gesendeten Daten lückenlos bei den Empfängern ankommen. Dies hat zur Entwicklung von spezialisierten Protokollen, aufbauend auf UDP, geführt, die die Dienste von UDP entsprechend anwendungsspezifisch erweitern. Die Protokollentwicklung ist jedoch aufwendig und gerade für komplexe Mehrpunktsituationen einer kleinen Gruppe von Spezialisten vorbehalten. So muss bei der Entwicklung vieler Anwendungen heute noch auf die Verwendung von mehreren Punkt-zu-Punkt-Verbindungen („TCP-Fanouts“) zurückgegriffen werden.

Auf dieser Basis ist bisher nur eine sehr begrenzte Anzahl an kommerziellen Produkten zur gesicherten Multicastübertragung entwickelt worden. Zu nennen wären hier die Produkte der Firmen Kencast, The Fantastic Corporation, Deuromedia und Starlight Networks. Aber auch diese Produkte bieten nur einen Bruchteil des für den universellen Einsatz von Multicast erforderlichen Funktionsumfanges und erfordern eine spezielle Implementierung auf die Gegebenheiten des jeweiligen Anwenders. So ist der Einsatz von IP-Multicast zur gesicherten Datenübertragung für eine sehr breite Gruppe an potentiellen Anwendern nur über Eigenentwicklung möglich.

2.2 ZIELSETZUNG

Als Pilotprojekt für die allgemeine Nutzung von gesicherter Datenübertragung über Multicast wird im Rahmen des Projektes MDiS durch die Firma Tellique Kommunikationstechnik GmbH und das Technologiezentrum Informatik der Universität Bremen eine Plattform zur Verfügung gestellt, die für verschiedenartige Anwendungen den Einsatz von zuverlässiger Multicast-Datenübertragung ermöglicht. Kern der Plattform ist das von der Tellique Kommunikationstechnik GmbH entwickelte Multicasttransportprotokoll MTP/SO. MTP/SO entspricht den Empfehlungen des Internet RFC 1301 und stellt quasi ein TCP-Pendant für Mehrpunktkommunikation dar.

Im Rahmen von MDiS werden zwei Möglichkeiten des Einsatzes von MTP/SO demonstriert, die beide als Beispielanwendung die Distribution von Dateien zum Ziel haben:

- Entwicklung von *mdist*
  Eines der unter Unix am häufigsten verwendeten Unicast-Dateiverteilsysteme ist *rdist*. Im Rahmen von MDiS wird *rdist* als Basis für die Entwicklung eines Multicast-Dateiverteilsystem unter Nutzung von MTP/SO verwendet. Damit wird ein Tool erstellt, durch dass sich MTP/SO einfach in bestehende, bewährte Systeme integrieren lässt. Nutzer von *rdist* können ohne Änderung ihrer bestehenden Systemkonfiguration Multicast nutzen.

- Piloteinsatz beim DWD
  Für den Piloteinsatz beim DWD wird zur Dateiverteilung die von Tellique auf Basis von MTP/SO entwickelte Datendistributionssoftware tq®-TELLICAST verwendet. Über tq®-TELLICAST werden zusätzliche Servicefunktionen zur Verfügung gestellt, die eine Einbindung in Datendistributionssysteme wie das AFD des Deutschen Wetterdienstes entscheidend erleichtern und zudem über Kompression, Verschlüsselung etc. die zentralen Vorteile von Multicast intensivieren.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Die aktuelle Implementierung von tq®-TELLICAST FileBroadcast wird von Tellique im Rahmen des Projektes in erweiterter Form bereitgestellt und zusätzlich den Einrichtungen des DFN-Vereins zur wissenschaftlichen Nutzung auf der Infrastruktur des Deutschen Wissenschaftsnetzes zur Verfügung gestellt. Auf Basis von tq®-TELLICAST wird durch das Projekt eine Plattform für Dateiübertragung als Beispielanwendung entwickelt, die es erlaubt, Dateigruppen automatisiert auf prinzipiell beliebig viele Empfängersysteme zu replizieren.
Diese Plattform wird vom Deutschen Wetterdienst (DWD) als Pilotnutzer im Rahmen von MDiS erprobt. Dazu wird das entwickelte Multicast-Distributions-System in die zentrale Infrastruktur des DWD eingebunden. Ziel ist unter anderem die Übertragung von im Meteorologischen Rechenzentrum in Offenbach erstellten Modelloutputs (Vorhersagedaten in hoher Auflösung) an 6 dezentrale Niederlassungen mit Regionalzentrale sowie in einem späteren Stadium die Verteilung von kleineren Datenbeständen an ca. 50 weitere Niederlassungen des DWD. Die Datenraten für die Verteilung an die dezentralen Niederlassungen innerhalb des Primärnetzes sollen dabei auf einen wesentlichen Teil der derzeit zur Verfügung stehenden 34 Mbit/s ansteigen. Angesteuert wird diese Übertragung durch das bereits bestehende Automated File Distributor (AFD) des DWD.

Pilotnutzer im Projekt ist der DWD, langfristig wird jedoch eine universellere Nutzung von reliable Multicast in den Netzen des DFN-Vereins angestrebt. Die Tellique Kommunikationstechnik GmbH und das Technologiezentrum Informatik der Universität Bremen können durch ihre bisherige direkte Mitarbeit in den Gremien des IETF dafür garantieren, dass deren derzeitige Forschungsergebnisse in das Design des beim DWD im MDiS-Projekt zu implementierenden Multicast-Systems einfließen. Es ist weiterhin anzustreben, die Ergebnisse des für den DWD durchgeführten Projektes auch in die Standardisierung einzubringen und damit zur weiteren universellen Verbreitung von Multicast in der gesicherten Datenübertragung beizutragen.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

3 HINTERGRUNDINFORMATIONEN

3.1 MULTICAST

3.1.1 Prinzip

Multicast ist die simultane Verteilung von Daten von einem Sender an mehrere definierte Empfänger. Die Daten brauchen dabei nicht an jeden Empfänger einzeln verschickt zu werden, sondern werden nur einmal in das Netz „eingestellt“. Netzkomponenten, wie beispielsweise Router, leiten die Daten an alle Empfänger weiter. Dabei leitet bei Bedarf ein Router beispielsweise Kopien eines eingehenden Paketes auf mehreren anderen Interfaces („in verschiedene Richtungen“) weiter. Auf den einzelnen Übertragungsstrecken werden die Pakete unabhängig von der Empfängeranzahl daher nur einmal übertragen (sofern keine Paketverluste auftreten). Dadurch wird die Netzauslastung, aber auch die Belastung des sendenden Servers (der sonst so viele Kopien versenden müsste, wie es Empfänger gibt) vermindert.

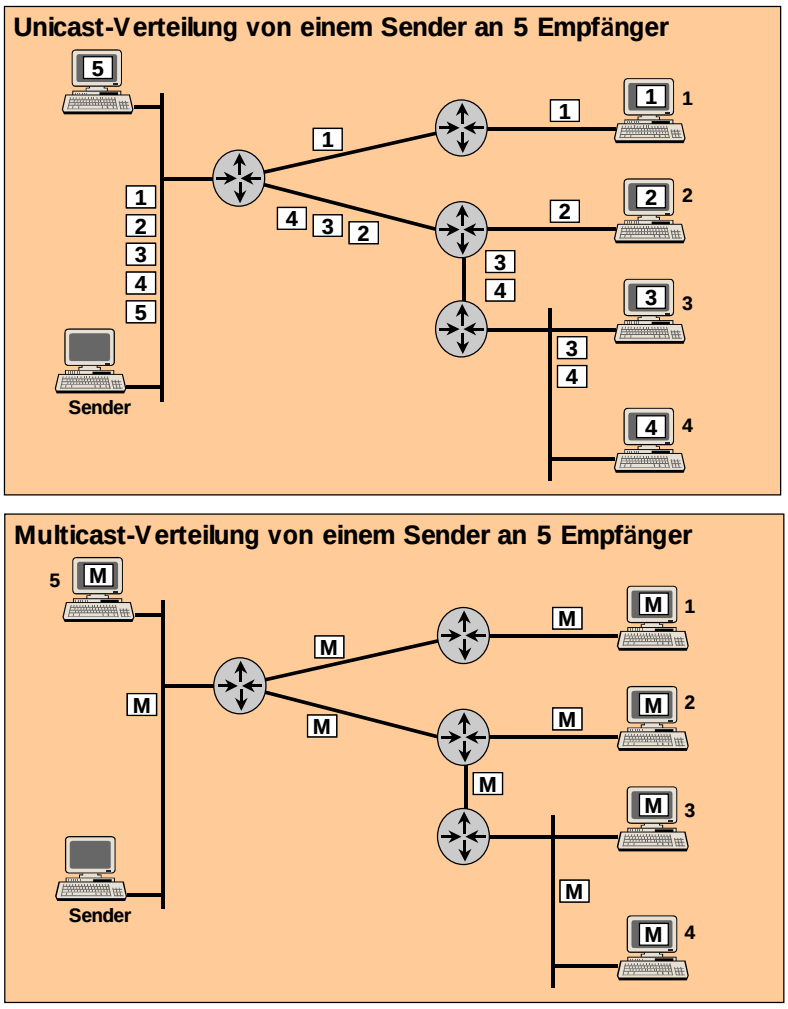

Abbildung 1: Unicast und Multicast im Vergleich

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Ein Endpunkt kann auf einer speziellen Multicastadresse Daten empfangen und senden. Hierfür ist im IP-Adressraum der Nummernraum 224.0.0.0 bis 239.255.255.255 vorgesehen („Klasse-D-Adressen“). Daten, die an eine Multicast-Adresse gesendet werden, werden von allen Rechnern empfangen, die auf dieser Adresse Daten empfangen wollen.

Häufig wird bei Multicast UDP als Transportprotokoll eingesetzt. TCP ist als Transportprotokoll für Multicast-Anwendungen nicht geeignet, da die Protokollmechanismen (z.B. die für Verbindungsauf- und -abbau und für Empfangsbestätigungen) eine einfache Zweierbeziehung voraussetzen. Eine typische Multicast-Anwendung ist z.B. die Übertragung von Audio- und Videoinformationen; hier wird auch im Unicast-Fall UDP eingesetzt.

Bei der Verteilung von Daten, bei denen es nicht auf die Echtzeitfähigkeit ankommt, sondern auf dem Empfang aller Daten in der richtigen Reihenfolge, muss ein anderes Transportprotokoll als UDP eingesetzt werden (oder auch der Dienst von UDP, Fehlererkennung und Anwendungsadressierung, durch ein weiteres Protokoll ergänzt werden). Hierfür kann das Multicast-Transportprotokoll MTP/SO verwendet werden.

Bei MTP/SO gibt es einen Master, der steuert, wann welcher Sender Daten senden darf. Empfängt der Master ein Paket, so gilt es als allgemein empfangen. Empfänger, die die Daten nicht erhalten haben können bei anderen Kommunikationspartnern diese Daten erneut anfordern. Durch Steuerpakete vom Master ist genau festgelegt, in welcher Reihenfolge Pakete an die einzelnen bei den Kommunikationspartnern laufenden Anwendungen ausgeliefert werden dürfen.

3.1.2 Vorteile

Durch die Nutzung von Multicast ergeben sich unter anderem die folgenden Vorteile:

- Verringerung der Netzlast und damit insbesondere im Weitverkehrsbereich oftmals erhebliche Einsparungen sowohl für die Anwender als auch für die Netzbetreiber.
- Deutliche Verringerung der Übertragungsverzögerung, da die Daten nur einmal übertragen werden, während der Sender bei Unicast-Übertragungen die Daten hintereinander an die einzelnen Empfänger senden müsste.
- Reduzierte Belastung der sendenden Rechner. Da die Daten nur einfach auszusenden sind, kann ein Rechner simultan auch Tausende von Empfängern erreichen – oftmals wird so eine breitbandige Versorgung großer Übertragungsgruppen überhaupt erst möglich.

Multicast entwickelt sich durch diese Vorteile zu einer Schlüsseltechnologie für die Datenübertragung des beginnenden 21. Jahrhunderts.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

3.1.3 Anwendungsgebiete

Die möglichen Anwendungsgebiete sind vielfältig und beinhalten z. B.

- Versorgung verteilter Organisationen mit aktuellen Informationsbeständen
- Realisierung von verteilten/replizierten Datenbanken
- Verteilte/replizierte Dateisysteme
- Softwareverteilung/Softwareupdates
- Groupware, Shared Applications, Audio-/Video-Konferenzen
- Übertragung von Daten vom „Information Service Provider“ an Abonnenten

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

3.2 STANDARDISIERUNG

Die massenhafte Verbreitung von Reliable-Multicast-Anwendungen im Internet setzt neben einer Multicast-Infrastruktur standardisierte Protokolle voraus. Dabei sind Normen vor allem in zwei Bereichen erforderlich:

- Transportprotokolle für Reliable Multicast
- Geeignete Sicherheitsstandards für große Gruppen

In beiden Bereichen stellte sich Mitte der 1990er Jahre schnell heraus, daß vor Beginn einer Standardisierung zunächst offene Forschungsfragen zu lösen waren. Die Aktivitäten wurden daher nicht sofort in die IETF getragen, sondern erst einmal in der *Internet Research Task Force* (IRTF) bearbeitet, bis sich jeweils so etwas wie ein Grundkonsens herausschälte. Abbildung [[[2]]] zeigt dies im Überblick.

Die Arbeiten im Bereich des Transportprotokolls in der IRTF bezogen sich zunächst auf die Entwicklung einer gemeinsamen Taxonomie, um überhaupt einmal der möglichen Vielfalt der Protokolle Herr zu werden. Später konzentrierte sich die *Reliable Multicast Research Group* (RMRG) auf für die Staukontrolle *(congestion control)* von Multicast-Strömen geeignete Verfahren sowie die Mindestanforderungen, die an solche Verfahren zu stellen sind *(Fairness)*.

	IRTF	IETF
Reliable Multicast	**RMRG** U early work U congestion control	**RMT** U bulk data transfer – unidirectional – bidirectional t NORM t TRACK
Multicast Security	**SMUG** U Security setup (LKH) –GDOI –GSAKMP U Source authentication	**MSEC** U Take up and standardize SMUG work

Abbildung 2: IRTF- und IETF-Aktivitäten in Transport und Security

Diese Arbeiten waren Ende 1999 so weit fortgeschritten, dass am 16.12.1999 die *reliable multicast transport* WG (RMT) in der IETF eingerichtet werden konnte. Dieser Arbeitsgruppe wurde jedoch eine klare Beschränkung auf den Bereich des Massendatentransports *(bulk transfer)* mit auf den Weg gegeben, da man die Lösung des Staukontrollproblems für den allgemeineren Fall noch nicht absehen konnte. Damit konnte diese

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Arbeitsgruppe mit ihren Protokollen nicht den Grad an Flexibilität erreichen, den MTP/SO auszeichnet.

RMT entschied sich wegen der immer noch hohen Zahl von möglichen technischen Alternativen nach einiger Diskussion dafür, die grundlegenden Mechanismen der Protokolle in Form von *Building Blocks* (BBs) zu entwickeln, von denen dann mehrere in *Protocol Instantiations* (PIs) referenziert und zu einem konkreten Protokoll ausentwickelt werden sollen.

	One-way	**Two-way**
Protocol BBs	u **FEC** u **LCT**	u **FEC** u **NORM** u **TREE**
Protocols (PIs)	u **ALC (LCT)**	u **NORM** –**(FEC, NORM)** u **TRACK** –**(FEC, NORM, TREE)**
Router Assist	u **(GRA for CC?)**	u **GRA**
Congestion Control	u **LCC**	u **TFMCC** u **PGMCC**

Abbildung 3: Das Arbeitsprogramm der IETF-Arbeitsgruppe RMT

Die Arbeiten der RMT-Gruppe spalteten sich schnell in eine auf Einwegkommunikation basierende Vorgehensweise *(Digital-Fountain*-Modell – es wird kontinuierlich gesendet, Empfänger können jederzeit beginnen zu empfangen, bis sie genug Bits haben, um den Inhalt zu rekonstruieren) und eine klassischere auf Zweiwegkommunikation basierende Vorgehensweise. Da für erstere noch nicht so viele technisch sinnvolle Protokolle auf dem Markt sind, konnte das Unternehmen *Digital Fountain* recht schnell Einigung über diese Protokolle erzielen; dementsprechend sind die Dokumente heute (Anfang 2002) bereits in der Nähe des *Working Group Last Call.*

Komplexer ist die Situation in der Zweiwegkommunikation. Hier muß zunächst zwischen der klassischen gruppenorientierten Kommunikation mit negativen Bestätigungen (*NACK-Oriented Reliable Multicast, NORM)* und komplexeren auf positiven Bestätigungen beruhenden baumbasierten Ansätzen (*Tree-based ACK, TRACK)* unterschieden werden. Nach extensiver Diskussion wurde klar, daß die TRACK-Lösung eine Obermenge der NORM-Lösung werden sollte. Drafts für beide Normen liegen vor, bedürfen aber noch weiterer Überarbeitung.

Eine weitere Schwierigkeit liegt in den Staukontrollalgorithmen für den Zweiwegefall. Neben dem in der RMRG entwickelten *TCP-friendly multicast congestion control* (TFMCC) trat mit der SIGCOMM 2000 PGMCC auf den Plan, eine schneller reagierende Variante der TFMCC-Grundidee, die aber entsprechend auch höhere Fluktuationen in der Datenrate mit sich bringt.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Zusammenfassend lässt sich sagen, dass (Stand Anfang 2002) eine Norm für den Zweiwegfall noch nicht absehbar ist. Wegen der schwierigen wirtschaftlichen Situation und der allgemeinen Desillusionierung bezüglich eines möglichen Internet-weiten Einsatzes von Multicast-Technologien wird die Normung gegenwärtig auf kleiner Flamme vorangetrieben. Für die absehbare Zukunft ist das auf RFC1301 basierende Protokoll MTP/SO deswegen auch weiterhin als lebensfähige Grundlage für die Entwicklung von Produkten im Bereich der Zweiwege-Multicastkommunikation anzusehen.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

4 MULTICAST-DATEIVERTEILUNG ÜBER *RDIST* MIT MTP/SO

Dieser Abschnitt beschreibt die Entwicklung eines Programms für die Verteilung von Daten per Multicast.

Es wäre ein einfaches Vorhaben gewesen, auf der Basis des ja bereits recht leistungsfähigen Multicast-Transportprotokolls MTP/SO eine simple Dateiverteilanwendung zu entwickeln. Es bietet sich aber an, Dateiverteilung per Multicast genau dort zu verwenden, wo bereits heute Dateiverteilung per Unicast im Einsatz ist. Das heute am häufigsten verwendete Programm für diesen Zweck ist das Programm *rdist*. Als Namen für dieses zu entwickelnde Programm wählen wir *mdist*, da *mdist* auf dem Quelltext von *rdist* basiert.

Im folgenden wird zunächst das Programm *rdist* beschrieben, welches als Basis für *mdist* verwendet wird. Daran anschließend erfolgt die Beschreibung des Programms *mdist*, dass über MTP/SO auf Grundlage von *rdist* die Verteilung von Dateien realisiert.

4.1 RDIST

Rdist wurde im Zusammenhang mit 4.3BSD entwickelt und ist deswegen in seiner ursprünglichen Form Bestandteil fast aller modernen UNIX-Systeme. Diese „klassische“ Variante von *rdist* hat allerdings einige schwerwiegende Mängel, so dass heute meist eine stark verbesserte Version im Einsatz ist.

Basis dieses Projektes zur Verteilung von Daten über Multicast ist das Programm *rdist* in der Version 6.1.5. Das Programm dient zur automatisierten Verteilung von Daten eines Quellrechners an mehrere Zielrechner. In einer Konfigurationsdatei können die Dateien, auch ganze Verzeichnisse, angegeben werden, die über das Netz verteilt werden sollen. Hierbei können auch Abweichungen auf den einzelnen Zielrechnern mit angegeben werden, wie z.B. ein anderes Zielverzeichnis oder das Überspringen bestimmter Dateien oder Verzeichnisse. Zusätzlich kann noch angegeben werden, ob neuere Dateien auf dem Zielrechner überschrieben und auf dem Quellrechner nicht mehr vorhandene Dateien auf dem Zielrechner gelöscht werden sollen.

Eine weitere wichtige Funktion von *rdist* ist das Ausführen von Programmen auf den Zielrechnern nach erfolgreicher Verteilung. Bei Verteilung von Programmen und Bibliotheken eines Linux-Systems kann mit dieser Funktion beispielsweise hinterher automatisch die Bibliotheken-Datenbank mit dem Aufruf *ldconfig* auf den aktuellen Stand gebracht werden.

Die Komplexität der Funktionalität von *rdist* spiegelt sich in der Konfigurationsdatei, dem Verteilungsmechanismus und dem Programmquelltext wider. An vielen Stellen sind Fallunterscheidungen und Funktionen mit Ausnahmeregelungen zu finden.

Beim Start des Programms werden *rdist*, die Konfigurationsdatei und eventuelle Änderungen zu dieser auf der Kommandozeile übergeben. Die Konfigurationsdatei wird mit einem mit Hilfe des Parser-Generators yacc (bzw. heute meist bison) erzeugten Parser in eine C-Datenstruktur übertra-

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

gen. Diese Datenstruktur wird nun weiter in die Befehle jedes Kommandos für jeden einzelnen Zielrechner zerlegt, immer unter Berücksichtigung der per Kommandozeile übergeben Änderungen. Am Ende dieser Verfeinerung steht eine Funktion, die genau ein Kommando auf einem Zielrechner ausführt.

Das Programm startet nun für jeden Zielrechner einen eigenen KindProzess[1], der einen Server[2] auf dem ihm zugewiesenen Zielrechner startet. Diese beiden Prozesse auf Quell- und Zielrechner tauschen nun Informationen und Daten für das jeweils zu bearbeitende Kommando aus.

Der Prozess auf dem Quellrechner überprüft anhand der Konfiguration, wie die Datei auf dem Zielrechner heißen soll und fordert Informationen über diese Datei an. Anhand dieser Informationen wird entschieden, ob die Datei übertragen werden muss oder nicht. Die Entscheidung, ob die Datei übertragen wird oder nicht, liegt beim Zielrechner, denn nur der *rdist*-Prozess auf dem Quellrechner weiß, ob neuere Versionen auf dem Zielrechner durch die Version auf dem Quellrechner überschrieben werden sollen.

Falls es sich bei dem Gegenstand der Verteilung um ein Verzeichnis handelt, wird anhand des Änderungsdatums überprüft, ob sich neuere Dateien in dem Verzeichnis befinden. Ist das Verzeichnis geändert worden, geht *rdist* rekursiv alle Dateien innerhalb dieses Verzeichnisses durch und überprüft wieder jede einzelne Datei.

Dieser Informationsaustausch ist durch ein auf ASCII basierendes Protokoll realisiert. Als Verbindung zwischen den Prozessen wird die Standardein- und -ausgabe verwendet. Ein Kommando besteht aus genau einer Zeile, die mit einem Code für die Art des Kommandos beginnt. Danach schickt der Server entweder eine positive (ACK) oder negative (REJ) Bestätigung oder angeforderte Informationen.

Entscheidet der Client, dass die Datei übertragen werden muss, so teilt er es dem Server mit und startet die Dateiübertragung.

Ist ein Kommando komplett abgearbeitet, überprüft der Client, ob noch andere Kommandos für diesen Server vorliegen. Ist dies der Fall, werden diese Kommandos ebenfalls von diesem Client-Server-Paar bearbeitet.

Ein großer Nachteil dieses von *rdist* verwendeten Dateiverteilverfahrens ist, dass für jeden Rechner die Datei separat auf dem Netz übertragen wird. Dieser Nachteil motiviert die Entwicklung eines auf Multicast basierendes Programms zur Dateiverteilung.

[1] Dieser KindProzess auf dem Quellrechner wird im folgenden Client genannt.
[2] Der Server auf dem Zielrechner ist der *rdist*-Daemon *rdistd*.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

4.2 DESIGN VON MDIST

Um die Funktionalität von *rdist* möglichst weitgehend zu erhalten, wurden nur wenige Änderungen am Original-Quelltext vorgenommen. Stattdessen wurden separate C-Module entwickelt, die spezielle Einsprungpunkte für die veränderte Aufgabe von *mdist* enthalten. Über das Setzen einer Variable wird der *mdist*-Teil im Programm aktiv. Durch diese „minimal invasive" Vorgehensweise kann *mdist* relativ leicht an eine neue Version von *rdist* angepaßt werden, sollte der Autor eine überarbeitete Version veröffentlichen[3].

Ein Hauptproblem war, die Punkte im Quelltext zu finden, an denen *mdist*-spezifische Funktionalität statt des bestehenden *rdist*-Codes ausgeführt werden muss. Das Programm *rdist* ist stark auf die Unterschiedlichkeit der Zielrechner ausgerichtet. Ein Client überträgt alle jeweils für ein spezifisches Ziel erforderlichen Daten. Bei Multicast sind aber die einzelnen Zielrechner bei der Datenübertragung nicht unterscheidbar, jeder empfängt die Daten und muss selbständig entscheiden, ob es sie annehmen muss oder nicht. Daher wurden zwei grundlegende Design-Änderungen vorgenommen:

- Die Intelligenz, welche Daten übertragen werden müssen, wurde vom Client zum Server verlagert. Hierzu benötigt der Server Informationen über die Kommandos, die Ausnahmeregelungen, Namen auf dem Quell- und Zielrechner und die Parameter auf der Kommandozeile für eventuelle Änderungen.
- Die Übertragung der Daten wird komplett am Ende durchgeführt, nachdem alle Server ihre Liste der Kommandos erhalten haben.

Kurz bevor die Aushandlung beginnt, welche Daten übertragen werden sollen, wird in den *mdist*-Teil des Programmes gesprungen. Statt mit dem Zielrechner jede einzelne Datei auszuhandeln und dann gleich zu übertragen, werden dem *mdist*-Prozess auf dem Zielrechner die Konfigurationsinformationen über diese Aufgabe übermittelt. Dann wird diese Aufgabe unterbrochen und die Konfigurationsdatei weiter bearbeitet.

Ist die Konfigurationsdatei komplett abgearbeitet worden, beginnt die Aushandlung, welche Daten übertragen werden sollen und danach die eigentliche Übertragung. Der Multicast-Teil von *mdist* besteht aus drei Phasen: die Übertragung der Dateiinformationen vom Client an den Server, die Antwort der Server mit der Liste der von ihnen benötigten Dateien und die eigentliche Datenübertragung aller benötigten Dateien vom Client zu den Servern.

4.2.1 Übertragung der Datei-Informationen

Der Client ermittelt eine Liste aller in der Konfiguration angegeben Dateien und Verzeichnisse. Die Liste besteht nur aus den Namen der Dateien auf dem Quellrechner; die Umsetzung auf die Dateinamen auf den Zielrech-

[3] Dies ist natürlich nur in begrenztem Umfang möglich. Substantielle Änderungen am Design von *rdist* erfordern wahrscheinlich ebenfalls eine Überarbeitung von *mdist*, da *mdist* auf die internen Datenstrukturen von *rdist* zurückgreift. Glücklicherweise kann *rdist* heute als weitgehend stabiles Programm angesehen werden.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

nern können diese anhand der per Unicast ausgetauschten Informationen selbständig ermitteln.

Zu jeder Datei und jedem Verzeichnis werden folgende Daten ermittelt und übertragen:

- Name auf dem Quellrechner
- Typ: Verzeichnis, Link oder reguläre Datei
- Dateigröße
- Änderungsdatum
- Rechte
- Besitzer, sowohl der Name als auch die systeminterne Kennziffer (UID)
- Gruppe, ebenfalls der Name und die Kennziffer

Die Übertragung der Namen und Kennziffern bei Besitzer und Gruppe sind nötig, da unter Umständen bei einigen Zielrechnern die Kennziffer und bei anderen der Name erhalten bleiben soll, falls diese auf dem Zielsystem nicht übereinstimmen.

Auf Serverseite werden diese Informationen der anfangs übertragenen Konfiguration zugeordnet bzw. verworfen, wenn die Datei auf diesem Server nicht benötigt wird. Auf dem Server nicht existierende Verzeichnisse werden in diesem Schritt angelegt, da an dieser Stelle schon alle relevanten Informationen ausgetauscht wurden.

Rdist ist in der Lage, Softlinks bei der Übertragung aufzulösen. Dadurch können Softlinks, die auf Dateien verweisen, die außerhalb des zu verteilenden Verzeichnisses liegen, ebenfalls mit übertragen werden. Es muss allerdings darauf geachtet werden, dass durch die Softlinks keine Schleifen bei Verweisen auf Verzeichnisse entstehen. Diese Funktionalität ist in *mdist* nicht mehr vorhanden. Dies ergibt sich aus dieser Art des Informationsaustauschs. Bei der Übertragung dieser Informationen werden erst alle Dateien an alle Server geschickt, bevor diese antworten, da es ansonsten zu unnötigen Verzögerungen kommt, wenn erst alle Antworten jeder Datei abgewartet werden müssen. Um Schleifen bei Softlinks auf Verzeichnisse zu vermeiden und da es unter Umständen sowohl Server gibt, die Softlinks aufgelöst haben wollen als auch solche, für die dies nicht vorgenommen werden soll, wurde auf diese nicht sehr oft benötigte Funktionalität verzichtet, da sie das Protokoll sehr verkompliziert und verlangsamt hätte.

4.2.2 Benötigte Dateien der Server

Nach der Übertragung aller Dateiinformationen wird der Server aktiv und muss die Liste der von ihm benötigten Dateien zurückschicken. Anhand der Konfiguration und der möglichen Dateinamen des Quellrechners werden die aus Sicht der Zielrechner zu verwendenden Dateinamen ermittelt[4].

[4] Je nach Komplexität der Konfiguration kann eine Datei des Quellrechners an mehreren Stellen beim Ziel zu speichern sein.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Nachdem diese Zuordnung durchgeführt ist, wird ermittelt, ob diese Datei neu übertragen werden muss. Hierbei werden vier Überprüfungen durchgeführt:

Die Konfiguration kann eine Liste von auszuschließenden Dateien und regulären Ausdrücken über den Dateinamen enthalten. Ist der Dateiname in dieser Liste, dann wird keine weitere Überprüfung durchgeführt, sondern die Datei als nicht benötigt markiert.

Handelt es sich um ein Verzeichnis, das noch nicht existiert, so wird der Dateiname in die Liste der anzufordernden Dateien aufgenommen. Der Client weiß bei Anforderung eines Verzeichnisses, dass er alle Dateien und Unterverzeichnisse dieses Verzeichnisses ebenfalls übertragen muss.

Ist die Datei innerhalb eines noch nicht existierenden Verzeichnisses oder ist sie die selbe Quelldatei wie eine schon als benötigt markierte Datei, so wird die Datei in die Liste der zu übertragenen Dateien aufgenommen, aber nicht in die Liste der Anforderungen, die an den Client geschickt werden.

Schließlich werden die weiteren Informationen über die Quell- und die Zieldatei ausgewertet. Hierzu werden Größe, Änderungsdatum und die Optionen der Konfiguration überprüft. Falls die Datei übertragen werden muss, wird sie in die Liste der benötigten und anzufordernden Dateien aufgenommen.

Die so erstellte Liste der Dateien und Verzeichnisse des Clients wird nun an diesen zurückgesendet und die eigentliche Datenübertragung kann starten.

4.2.3 Datenübertragung

Nachdem alle Server die Liste der von ihnen benötigten Dateien übertragen haben, beginnt die Übertragung aller benötigten Dateien selbst. Am Anfang wird der Dateiname auf dem Quellrechner und die Größe übertragen. Danach wird diese Datei übertragen und dann die nächste Datei bearbeitet.

Jeder Zielrechner empfängt alle Dateien. Er muss dann selbständig entscheiden, ob er sie speichern muss oder nicht. Die Informationen über die Rechte, die diese Datei bekommt, erhält der Server aus den Informationen, die in Phase 1 ausgetauscht wurden.

4.2.4 Verwendung von rdist zur abschließenden Fehlerbehebung

Während der Multicast-Übertragung kann durch längere Ausfälle der Multicast-Konnektivität der Fall auftreten, dass ein oder mehrere Zielrechner die Übertragung nicht komplett erhalten. Sie ziehen sich bei Problemen aus der Multicast-Übertragung zurück und melden dem Client, dass sie die Daten nicht empfangen können.

Nachdem die Multicast-Übertragung abgeschlossen ist, wird die Variable, die den *mdist*-Teil des Programms aktiviert, auf die Verwendung des *rdist*-Protokolls zurückgesetzt. Danach wird das Programm mit den nicht abgeschlossenen Operationen als *rdist* noch einmal gestartet; damit werden etwaige Lücken der Multicast-Übertragung automatisch behoben.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

4.2.5 Programmdokumentation zu mdist

Der *mdist*-Teil des Programms besteht aus vier C-Modulen: *mtp*, *mclient*, *mserver* und *mparser*.

Der mtp-Teil sorgt für die Multicast-Übertragung auf Basis der Bibliothek mtpso. Er enthält Hilfsfunktionen für die Initialisierung eines mtpso-Multicast-Sockets und für das Senden und Empfangen von Daten. Diese Funktionen werden vom *mserver* und dem *mclient* verwendet.

Die Module *mclient* und *mserver* enthalten den eigentlichen *mdist*-Teil des Programms. Sie enthalten die Funktionen, die vom Original-*rdist* angesprungen werden. Für die Parsierung der Datenströme für den Informationsaustausch wird das Objekt *mparser* verwendet. Es enthält Funktionen, um interne Datenstrukturen in Zeichenketten und zurück zu übersetzen.

4.2.6 Protokollbeschreibung

Die Unicast- und die verschiedenen Phasen der Multicast-Übertragung setzen verschiedene Protokolle für die Datenübertragung ein. Die Übertragung der Konfiguration und der Dateiinformationen ist ein auf ASCII aufsetzendes, zeilenorientiertes Protokoll. Die verschiedenen Informationen werden durch Leerzeichen getrennt, Zahlen in menschenlesbarer Form übertragen. Zeichenketten, die wiederum Leerzeichen enthalten können, enthalten eine zusätzliche Längenangabe, um sie wieder korrekt extrahieren zu können.

Jede Zeichenkette beginnt wie bei *rdist* auch mit einem Zeichen, welches die Art der Zeile beschreibt. Die einzelnen Codes sind in der nachfolgenden Tabelle aufgelistet.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

4.2.7 Bewertung des Ergebnisses

Mdist ist ein auf der Basis von *rdist* entwickeltes System zur Dateiverteilung per Multicast. Der Ansatz, nicht ein völlig neues Tool zu entwickeln, sondern auf *rdist* einzusetzen, bietet Anwendern von *rdist* eine sehr einfache Migrationsstrategie in Richtung Nutzung von Multicast zur Dateiverteilung. Projekte, die z.B. im Wissenschaftsnetz des DFN größere Dateien regelmäßig verteilen wollen, erhalten mit *mdist* ein Tool, das die benötigten Bandbreiten deutlich reduzieren kann.

Rdist ist natürlich nicht das einzige Tool zur Dateiverteilung, das als Basis für die Entwicklung von *mdist* hätte dienen können. Ein anderes heute weit verbreitetes Programm zur Dateiübertragung ist *rsync*. Statt *rdist* hätte auch *rsync* auf Multicast-Verteilung angepaßt werden können. Im Gegensatz zu *rdist* ist *rsync* aber allein auf die Verteilung ausgerichtet und hat weniger Sonderfunktionen für den Einsatz in größeren Netzen; so kann *rsync* beispielsweise nach erfolgreicher Verteilung nicht Programme zur endgültigen Installation der verteilten Dateien ausführen. Die grundsätzliche Idee von *rsync*, durch Überprüfung partieller Prüfsummen der Dateiinhalte Übertragungskapazität einzusparen (Tridgell, A. and Mackerras, P., „The rsync algorithm“. Technical Report TR-CS-96-05, Dept. Computer Science, ANU, 1996), bietet im Multicast-Fall auch weniger Vorteile.

Die Entscheidung, *rdist* als Basis für das Programm *mdist* zu verwenden, ist jedoch nicht nur mit Vorteilen verbunden. Einerseits wird in vielen Infrastrukturen heute *rdist* eingesetzt und dementsprechend ist in vielen Organisationen signifikanter Aufwand investiert worden in die Entwicklung von *rdist*-Konfigurationsdateien, die nun mit *mdist* einfach unverändert übernommen werden können. Die Funktionsvielfalt von *rdist* macht auch *mdist* zu einem mächtigen Werkzeug für die Verteilung von Daten.

Diese Funktionsvielfalt und die damit einhergehende Komplexität ist andererseits auch einer der größten Nachteile des entstandenen Systems. Der Quellcode von *rdist* ist an vielen Stellen sehr verwickelt und daher unübersichtlich. Da die Intelligenz bei *mdist* im Server und nicht wie bei *rdist* im Client steckt, mussten allerlei Funktionen neu implementiert werden. Durch die Komplexität ist es auch sehr schwierig, alle denkbaren Spezialfälle zu testen, weil sie abhängig von der jeweiligen Anwendung sind und der Phantasie der Anwender bei der Anwendung der *rdist*-Konfigurationssyntax kaum Grenzen gesetzt sind.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

5 MULTICAST-DATEIVERTEILUNG DURCH tq®-TELLICAST

5.1 PRODUKTAUFBAU tq®-TELLICAST

tq®-TELLICAST ist ein Multicastübertragungssystem, das auf MTP/SO als Multicast-Transportprotokoll basiert. Es besteht aus verschiedenen Modulen, die die Übertragung unterschiedlicher Inhalte ermöglichen:

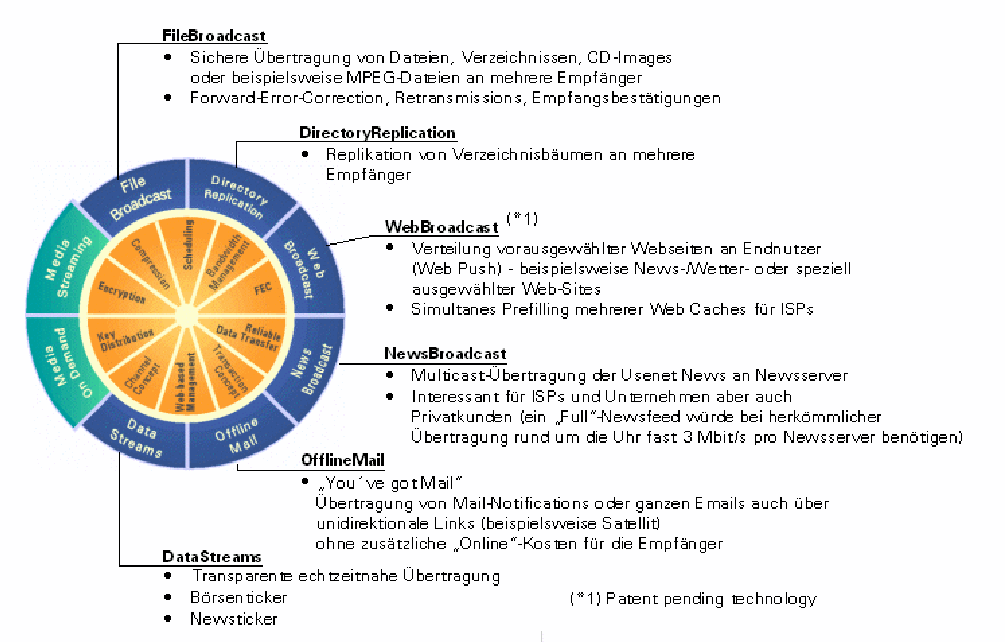

Abbildung 4: Produktmodule von tq®-TELLICAST

Der modulare Produktaufbau ermöglicht den Einsatz des Systems in einer Vielzahl unterschiedlicher Einsatzbereiche und zugleich den gezielten Einsatz auf die jeweilige Content-Art angepaßter Module für das jeweilige Projekt. Für den Einsatz im DWD wird die Funktionalität von tq®-TELLICAST auf das Modul FileBroadcast beschränkt.

Alle Module von tq®-TELLICAST weisen neben bzw. durch die Verwendung von MTP/SO die folgenden Sevicefunktionen auf, die zugleich die Basis für die Einbindung in das Dateiübertragungssystem des DWD darstellen:

- zeit- und prioritätenkontrollierte Steuerung durch einen zentralen Scheduler
- Übertragung in logischen Kanälen (Channel), die u. a. die Festlegung von maximalen Bandbreiten pro Channel erlauben
- zentrale Bandbreitensteuerung (Überwachung der Gesamtband-breite, Koordinierung paralleler Sendeaufträge)
- Datenkomprimierung „on the fly", d. h. während der Übertragung und somit ohne zusätzliche Verzögerung
- Datenverschlüsselung „on the fly"
- gesicherte Datenübertragung mit gezielten Retransmissions

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

- Überwachung durch ein webbasiertes Interface und Log-Dateien
- Ausgabe von für das Accounting relevanten Daten in spezifische Dateien

5.2 FUNKTIONSWEISE

Der tq®-TELLICAST Sender ermöglicht die Multicast-Übertragung von Dateien an mehrere Empfänger. Die Daten werden als Dateien, Dateiverzeichnisbäume oder – für Datenströme wie z. B. Börsenticker – als transparenter Datenstrom übermittelt.

Die Übertragung erfolgt auf einer per-Channel-Basis, d. h. der physikalische Kanal kann in viele logische Channel unterteilt werden. Für jeden Channel werden über Channel Files spezifische Parameter wie Bandbreite, Kompression etc. festgelegt. Die Channel Files enthalten auch eine Referenz auf ein oder mehrere Recipient Files. Die Recipient Files sind Empfängerlisten. Alle Empfänger in den durch das Channel File referenzierten Recipient Files können die Daten dieses Channels empfangen.

5.2.1 Initialisierung / Announcements

Über einen Announcement Channel informiert der Sender die Empfänger kontinuierlich über alle anstehenden Übertragungen. Dazu gehört beispielsweise die:

- Ankündigung aller Übertragungen vor dem Datenaustausch,
- Verteilung der Datenschlüssel für verschlüsselte Übertragungen und
- Auffordern der Empfänger, Bestätigungen für die erhaltenen Daten zu senden.

Insgesamt hat der Announcement-Channel die Aufgabe, die Menge der auf den Empfängern zu konfigurierenden Daten möglichst klein zu halten und die notwendigen Informationen stattdessen automatisch vom Sender zu empfangen.

5.2.2 Übertragung

Sender

Die Übertragung wird durch sogenannte Job Files angesteuert. Als Job wird die Übertragung einer bestimmten Datenmenge an eine spezifizierte Gruppe von Empfängern bezeichnet. Ein Job File definiert die Parameter, die für die Administration und Ausführung einer spezifischen Datenübertragung notwendig sind.

Der tq®-TELLICAST Sender liest kontinuierlich neue Job Files aus dem Verzeichnis „jobs/incoming“ ein. Ein neuer Job wird ausgewertet und kontrolliert, mit Systemparametern erweitert und in das Verzeichnis „jobs/scheduled“ verschoben.

Solange ein Job aktiv ist, verbleibt er im Verzeichnis „jobs/scheduled“. Der aktuelle Zustand der Jobs wird regelmäßig in den Job Files im Verzeichnis „jobs/scheduled“ zwischengespeichert. Dies ermöglicht die Fortführung der

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Übertragung nach einem Neustart des Systems an weitgehendst der selben Stelle, an der die Unterbrechung erfolgte.

Nachdem ein Job beendet wurde, wird das entsprechende Job File in das Verzeichnis „jobs/done“ verschoben, um das Ende der Übertragung anzuzeigen. Job Files im Verzeichnis „jobs/done“ werden nicht länger vom System verwendet und können entfernt oder geändert werden.

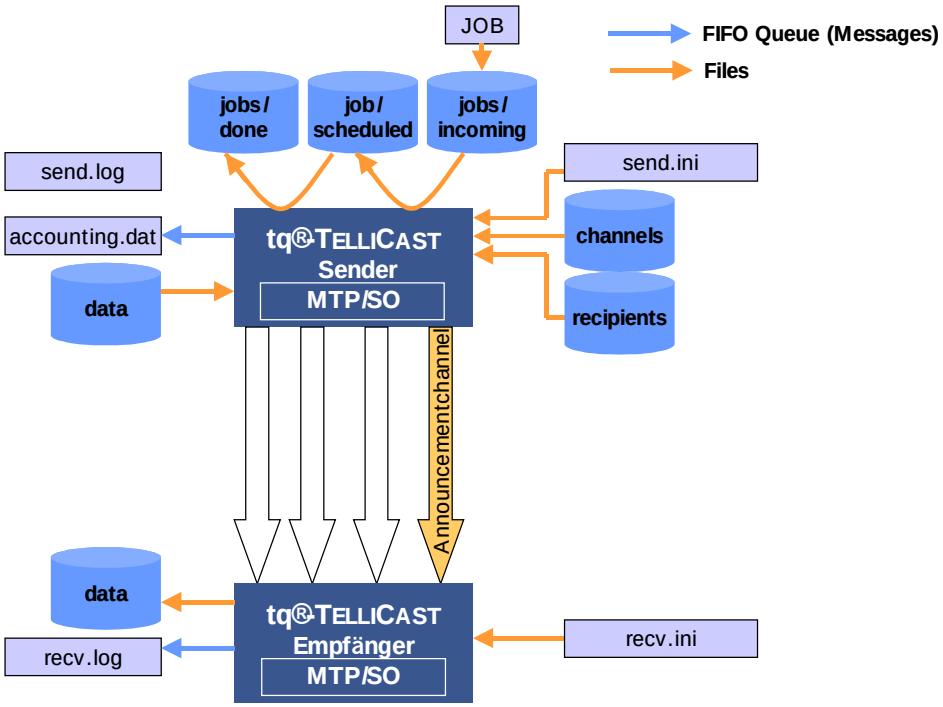

Abbildung 5: Dateienübertragung mit tq®-TELLICAST

Empfänger

Beim Systemstart öffnet der Empfänger den allgemeinen Announcement Channel und wartet auf die Verteilung von Schlüsseln (bei aktivierter Datenverschlüsselung), Übertragungsankündigungen und Empfangsaufforderungen. Erfolgt eine Empfangsaufforderung, so werden die Daten auf dem angekündigten Kanal empfangen.

Die empfangenen Dateien und Verzeichnisse werden unter dem selben Namen und unter den selben Unterverzeichnissen abgelegt, die auch der Sender benutzt, jedoch in Relation zu einem kanalspezifischen Startverzeichnis, das pro Empfänger gesondert konfiguriert werden kann.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6 ANWENDUNGSFALL DWD

Kernbestandteil des Projektes MDiS war die Realisierung von Multicast-Datenübertragung innerhalb der bisher über AFD erfolgenden Datenverteilung des DWD.

Dieses Kapitel gibt zunächst einen groben Überblick über die zugrunde liegende Kommunikationsinfrastruktur des DWD und die Anforderungen, die von Seiten des DWD an eine Multicast-Datenverteilung gestellt werden. Auf dieser Basis wird dann die Zielarchitektur vorgestellt.

6.1 DATENVERTEILUNG IM DWD

Im Folgenden wird ein Überblick über die Kommunikationsinfrastruktur und die Datenverteilung innerhalb des DWD gegeben, sofern diese für MDiS relevant sind.

6.1.1 Beteiligte Netzstrukturen

6.1.1.1 Primärnetz

Das Primärnetz verbindet ringförmig die Zentrale des DWDs in Offenbach mit den 6 dezentralen Niederlassungen mit Regionalzentrale in Essen, Hamburg, Potsdam, Leipzig, München und Stuttgart. Die Standleitungen des Primärnetzes haben im derzeitigen Ausbau eine Bandbreite von 34 MByte/s, von denen 8 Mbit/s zur Datenverteilung der angebundenen anderen Behörden des Bundesministeriums für Verkehr, Bau- und Wohnungswesen (BVBW) reserviert sind.

Den überwiegenden Anteil der Datenübertragung innerhalb des Primärnetzes macht die Datenversorgung der dezentralen Niederlassungen mit Regionalzentrale aus, wobei die Datenbanktransaktionen nur einen geringeren Anteil ausmachen, während die Dateiverteilung den größten Teil des Datenvolumens benötigt (im Endausbau bis 12 Gbyte Daten).

In Rückrichtung von den dezentralen Niederlassungen mit Regionalzentrale an die Zentrale in Offenbach erfolgt die Übertragung von Messdaten, z. B. der Radarstandorte. Die benötigte Bandbreite ist jedoch viel geringer als die für die Produktverteilung von der Zentrale an die Niederlassungen mit Regionalzentrale.

6.1.1.2 Sekundärnetz

Das Sekundärnetz verbindet weitere größere Standorte des DWDs über Standleitungen mit den dezentralen Niederlassungen mit Regionalzentrale. Dies sind vor allem Flugwetterwarten, Radarstandorte und Observatorien. Die Bandbreite dieser Verbindungen ist mit 64 kbit/s geringer als die des Primärnetzes. Es erfolgt je nach der Auslegung des Standortes eine Übertragung von der Zentrale an die Standorte und/oder Übermittlung von Messdaten an die Zentrale.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

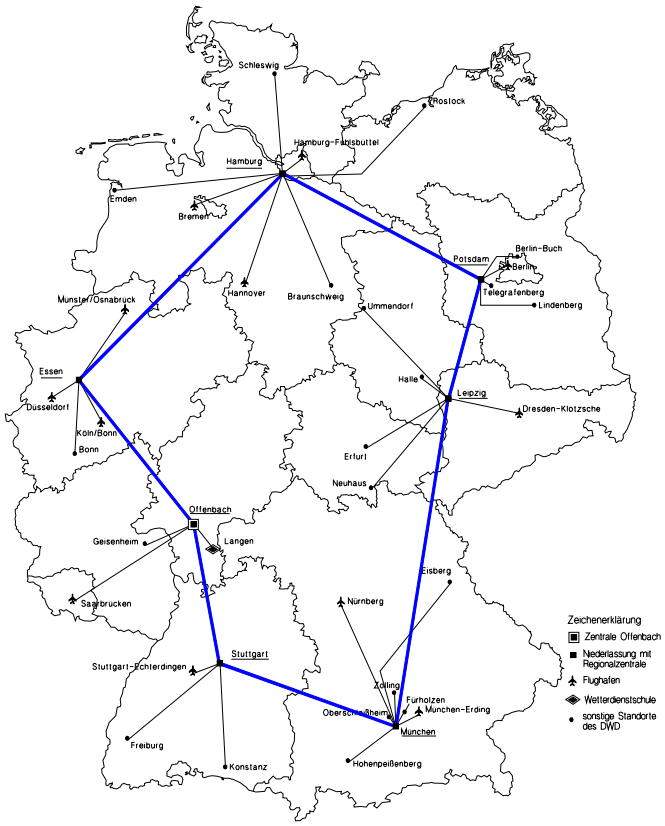

Abbildung 6: Primärnetz und Sekundärnetz des DWDs

6.1.2 Datendistribution über AFD

AFD ist ein vom DWD entwickeltes Dateidistributionssystem, das sowohl die Datenübertragung innerhalb des Primär- und Sekundärnetzes als auch Teile der Kundenversorgung und des internationalen Austausches von meteorologischen Daten steuert und durchführt.

Über den Automated File Distributor (AFD) werden die Produkte von der Zentrale in Offenbach aus an alle Niederlassungen verschickt. Dabei beachtet AFD die physikalische Ringstruktur des Primärnetzes, indem die Daten von Regionalzentrale zu Regionalzentrale weitergereicht werden. So sendet die Zentrale die Daten nur an die beiden nächstgelegenen Regionalzentralen. In einer Regionalzentrale werden die Daten sowohl regional verteilt als auch sofort an die auf dem Ring nachfolgende Regionalzentrale gesendet. Dadurch erfolgt die Verteilung der Daten in die DWD-Fläche über zwei „Halbringe“, nämlich über Essen und Hamburg nach Potsdam und über Stuttgart und München nach Leipzig.

Bei dieser Verteilstruktur ergeben sich Probleme bei Ausfall einer Regionalzentrale, da dann die auf dem Ring nachfolgenden Regionalzentralen nicht versorgt werden. In diesem Fall werden die Daten vom Endpunkt eines „Halbringes“ an die nachfolgende Regionalzentrale, also z. B. von Potsdam nach Leipzig, gesendet. Dies ist durch einfache Operatoranweisungen am jeweiligen AFD zu erreichen.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Jede Niederlassung besitzt einen Backup-Rechner für den Rechner, auf dem die Daten eingehen. Die Zugangsdaten und Zielverzeichnisse dieses Rechners sind mit denen des eigentlichen Empfangsrechners identisch. Der Backup-Rechner wird von AFD versorgt, sobald der ursprüngliche Empfangsrechner ausfällt.

6.1.3 ISDN-Failover

Bei Ausfall von Standleitungen des ATM-basierten AFD-Systems werden die Daten an die betroffenen Rechner über ISDN übertragen. Die ISDN-Übertragung wird mit einer sehr viel geringeren Bandbreite als die primäre AFD-Übertragung realisiert (unter Einsatz von Kanalbündelung bis zu 2Mbit/s zu einem Standort).

Die Datenserver empfangen nicht nur über das interne ATM-Interface direkt Daten aus dem Primärnetz, sondern sind zusätzlich über ein Ethernet-Interface an das lokale LAN angebunden, über das sie u.a. auch die im Failover-Fall über ISDN übertragenen Daten empfangen können.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.2 GEPLANTE DATENÜBERTRAGUNG ÜBER MULTICAST

6.2.1 Vorgesehener Einsatzbereich

Multicast soll zunächst nur für die DWD-interne Datenübertragung eingesetzt werden. Für die Übermittlung von Daten an Kunden kann dieses Verfahren ggf. in einer späteren Ausbaustufe und evtl. kombiniert mit Satellitendiensten erweitert eingesetzt werden.

Im Rahmen von MDiS soll die Multicast-Datenübertragung im Primärnetz realisiert werden. Das Sekundärnetz wird im Rahmen des Projektes nur versuchsweise und voraussichtlich zunächst nur an einem Standort durchgeführt, mit dem Ziel, die Einsetzbarkeit in diesem Bereich sicherzustellen.

Damit ergeben sich folgende Einsatzbereiche für Multicast:

- Übertragung von Daten von der Zentrale in Offenbach an die Daten-Server in den Niederlassungen mit Regionalzentrale.
- Übertragung der von den Stationen des Sekundärnetzes an den Regionalzentralen eingehenden Daten von den Regionalzentralen an die Zentrale in Offenbach.

Die Übertragung der Daten erfolgt auf Sendeseite über Vermittlung durch AFD, d. h. es wird eine Datei-/Verzeichnis-basierte Schnittstelle definiert, über die AFD einzelne Übertragungsaufträge an das Multicastsystem weitergibt. Dabei ist es alternativ möglich, auch direkt von anderen Systemen (außer AFD) Daten zur Übertragung an das Multicastsystem weiterzugeben.

Auf Empfängerseite sind zunächst die Daten-Server als Empfänger vorgesehen. In einem späteren Schritt (außerhalb des MDiS-Projektes) werden direkt die Applikationsrechner als zusätzliche Empfänger eingerichtet. Die Übertragung erfolgt parallel an beide Daten-Server, da dies über Multicast keine Erhöhung der Bandbreite gegenüber der Übermittlung von Daten an einen einzelnen Rechner darstellt und so eine hohe Ausfallsicherung erreicht werden kann.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.2.2 Anforderungen

Für die Realisierung des Multicastsystems ergeben sich folgende grundlegende Anforderungen von Seiten des DWDs:

- Jede Regionalzentrale muss als Empfänger und als Sender fungieren können, d. h. das System muss mehrere Sender zulassen, die zentral koordiniert werden.
- Eine Übertragung muss parallel an mehrere Rechner in den Regionalzentralen erfolgen können.
- Auf den Zielrechnern werden die Daten vom Multicastsystem in einem pro Übertragungschannel definierbaren Zielverzeichnis angelegt. Eine weitere Unterverteilung erfolgt bei Bedarf durch AFD.
- Die Übertragung der Daten muss bandbreitengesteuert und prioritätengesteuert erfolgen können.
- Die ISDN-Übertragung als Backup muss realisierbar sein.
- Das Multicastsystem soll eine Schnittstelle zum Abfragen von Zustandsinformationen bieten. Ziel ist es, über ein Graphical User Interface (GUI) die Sendezustände der beteiligten Rechner abfragen zu können. Das GUI wird vom DWD ähnlich dem AFD GUI entwickelt.
- Die unterschiedlichen Bandbreiten für das Sekundärnetz und das Primärnetz sowie für eine Failover-Übertragung über ISDN müssen durch die Definition mehrerer Channel (Multicast-Übertragungsgruppen) und potentiell unterschiedlicher Bandbreiten pro Übertragungsgruppe abgebildet werden können.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.3 ZIELARCHITEKTUR

6.3.1 Integration von tq®-TELLICAST und AFD

Ziel des zu entwickelnden Multicast-Distributionssystems ist es, angesteuert durch AFD, einen gesicherten Multicast-Übertragungsdienst für mehrere Gruppen anzubieten und dadurch neben z. B. Unicast-FTP einen weiteren Übertragungspfad zur Verfügung zu stellen.

MTP/SO setzt dabei auf UDP auf. Zwischen dem Service-Layer von MTP/SO, tq®-TELLICAST, und AFD wird zur Schnittstellensteuerung ein Adaptation-Layer implementiert, der im folgenden als Multicast-Distribution Layer (MD) bezeichnet wird.

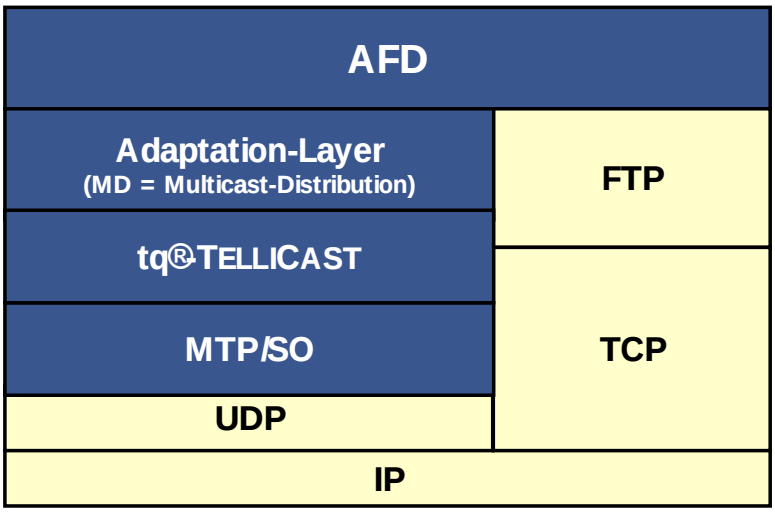

Abbildung 7: Schichtenmodell zur Einbindung von Multicast in AFD

Die folgenden Ausführungen beschreiben darauf aufbauend die Übergabe der Daten vom AFD-System über MD an tq®-TELLICAST und die Rückmeldung des Sendeerfolgs von tq®-TELLICAST über MD an das zu entwickelnde MDiS-GUI.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.3.2 Schnittstelle zur Datenübertragung: MD

Die Schnittstelle zwischen AFD und tq®-TELLICAST ist ein neu zu implementierender Prozess „Multicast Distribution Layer“ (MD) und ein bereits nach Channeln organisierter Dateibaum.

6.3.2.1 Schnittstelle auf Senderseite

AFD (oder zukünftig auch andere Systeme) stellen die zu übertragenden Dateien auf Sendeseite in Channel Directories ein. Für jeden definierten Übertragungschannel wird hierfür auf Sendeseite ein Channel Directory vorgesehen (siehe Abbildung 8).

Der Prozess MD scannt die Channel Directories (einschließlich enthaltener Dateien und Unterverzeichnisse) kontinuierlich. Alle dort neu erkannten oder modifizierten Dateien werden zu Sendeaufträgen (Jobs) zusammengefasst und über Job-Files zur eigentlichen Multicast-Übertragung an tq®-TELLICAST übergeben.

Entsprechend der genutzten Channel-Directories überträgt tq®-TELLICAST die Dateien auf dem jeweiligen Übertragungs-Channel.

Die GUI (oder alternativ auch ein anderer Prozess) verbinden sich über einen TCP-Server-Port mit dem MD und rufen die Daten zum aktuellen Stand der Übertragungen über ein ASCII-basiertes Protokoll ab.

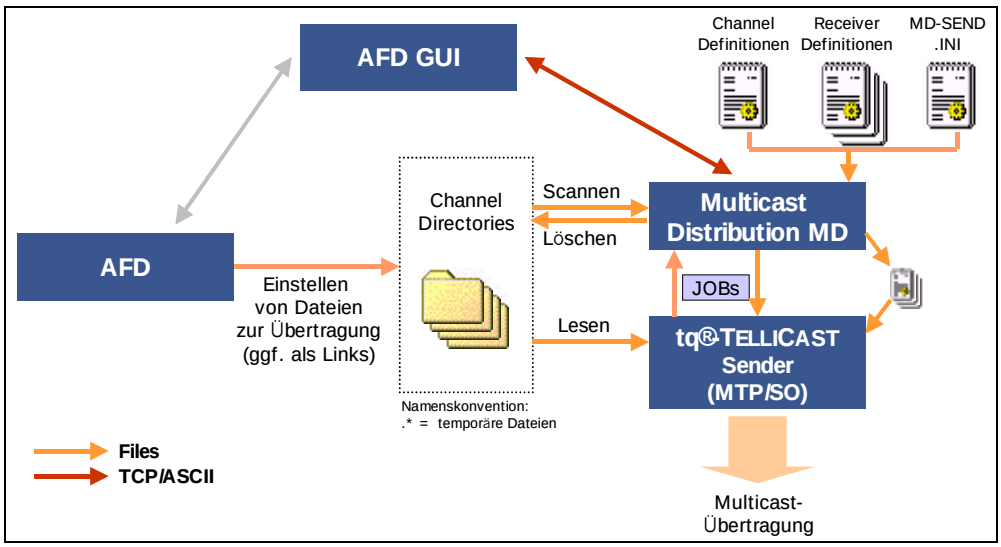

Abbildung 8: Schnittstellen zwischen AFD und tq®-TELLICAST Sender

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.3.2.2 Schnittstelle auf Empfängerseite

Empfängerseitig empfängt tq®-TELLICAST den Multicast-Announcement Channel sowie die für diesen Empfänger konfigurierten Multicast-Data-Channel.

Welche Channel von einem Receiver empfangen werden, kann vom Administrator durch eine Receiver-Channel-Definitionsdatei festgelegt werden. In dieser Definitionsdatei ist neben den zu empfangenden Channeln u.a. auch pro Channel ein Verzeichnis konfiguriert, in das die auf diesem Channel empfangenen Dateien abgelegt werden sollen.

Der Prozess MD hat auf Empfangsseite vorwiegend die Aufgabe, die in einem MDiS-spezifischen Format angegebenen Channel-Definitionen sowie die Parameter einer zentralen md-recv.ini Datei in die tq®-TELLICAST spezifischen Konfigurationsdateien umzuwandeln. Dabei können auf Empfangsseite, ebenso wie auf Senderseite, die Angaben in der Channel Definitionsdatei überwiegend auch zur Laufzeit ohne einen Neustart der MD/ tq®-TELLICAST Prozesse vom Operator geändert und neu übergeben werden.

tq®-TELLICAST legt die empfangenen Dateien direkt in dem für den jeweiligen Channel angegebenen Verzeichnis ab. In diesem Verzeichnis stehen sie für andere Anwendungen zur Nutzung oder AFD zur weiteren Verteilung zur Verfügung. Ein Löschen dieser Dateien erfolgt durch AFD oder zukünftig ggf. auch alternative Systeme.

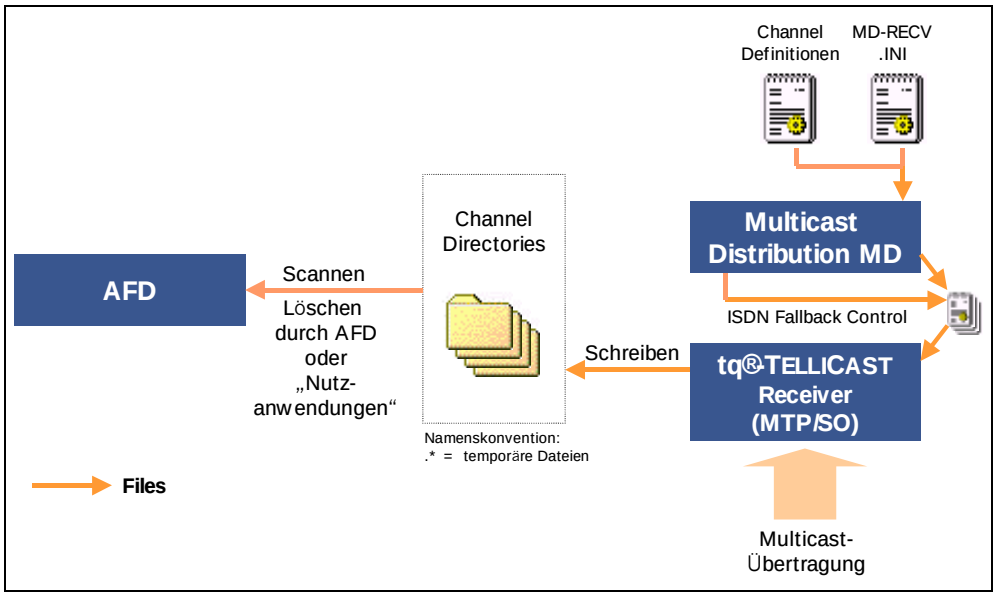

Abbildung 9: Schnittstellen zwischen AFD und tq®-TELLICAST Receiver

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.4 GESICHERTE DATENÜBERTRAGUNG

Die gesicherte Übertragung der Daten durch MDiS ist, neben der gleichzeitigen Übertragung an Daten-Server und Backup-Rechner, durch vier Mechanismen gewährleistet:

- Forward Error Correction (FEC),
- Anforderung von Datenpaketen über „Negative Acknowledgements“ (NAK),
- Retransmission auf Grundlage von Acknowledgements und
- das ISDN-Backup-System des DWD.

6.4.1 Forward Error Correction

FEC gewährleistet den Ausgleich von Übertragungslücken auch bei Fehlen eines Rückkanals, wie z. B. bei Satellitenübertragungen. Vorteil im zunächst ausschließlich terrestrischen Einsatz des DWD ist, dass Übertragungsfehler (fehlende Pakete) durch den Einsatz von FEC behoben werden können, ohne dass einer oder sogar mehrere der Empfänger explizite NAKs (negative Empfangsbestätigungen) an den Sender zurücksenden müssen.

Der Sender generiert auf Ebene von MTP/SO FEC Redundanzpakete aus den Daten mehrerer Datenpakete, die zusammen mit den eigentlichen Daten übertragen werden, und aus denen sich fehlende Datenpakete errechnen lassen.

Die Generierung der Redundanzpakete erfolgt „on the fly“ und unmittelbar während der Übertragung.

6.4.2 Negative Acknowledgements (NAK)

Wenn einzelne Datenpakete nicht empfangen werden, fordert der Empfänger das Paket durch Senden eines NAKs neu an. Das fehlende Paket wird daraufhin noch einmal übertragen. Diese Kontrolle des Datenempfangs wird auf der Ebene von MTP/SO realisiert.

6.4.2.1 Bandbreitenkontrolle für NAK

Der Verbrauch von Bandbreite durch NAK-gesteuerte Paketanforderungen ist schwer vorhersehbar. Starke Empfangsstörungen einzelner Empfänger können zu einer hohen Belastung der Bandbreite des Channels führen, vor allem wenn sehr viele Empfänger zum Empfang eines Channels berechtigt sind.

Für diese großen Empfängergruppen kann die Belastung des Senders durch die Beantwortung von NAK-Anfragen durch Repetitoren verringert werden. Als Repetitoren werden Empfänger bezeichnet, die nicht nur die Daten empfangen, sondern außerdem gegenüber anderen Empfängern als Sender fungieren und deren NAK beantworten. Die Bandbreite zum Sender hin wird damit entlastet, da der Sender nur noch die NAK des Repetitors erhält, während die NAK aller Empfänger im Einzugsbereich des Repetitors von diesem beantwortet werden.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Die NAKs werden von den Empfängern, die dem Repetitor zugeordnet sind, zunächst mit einer geringen TTL (time to live; Anzahl der Router, die ein Datenpaket passieren kann) gesendet. Dadurch ist gewährleistet, dass sie nur bis zum Repetitor übertragen werden und den Sender nicht mehr erreichen. Nur in dem Fall, dass der Repetitor ausfällt und die fehlenden Datenpakete nicht an den Empfänger überträgt, sendet dieser die NAKs noch einmal mit höherer TTL aus, so dass eine Anfrage beim Sender erfolgt

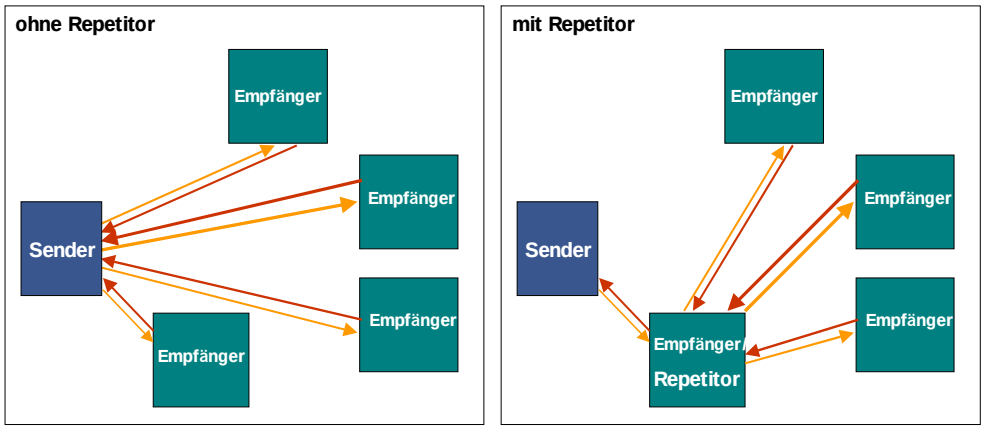

Abbildung 10: NAK-Retransmissions ohne und mit Repetitor

In die Regulation der Gesamtbandbreite für die Übertragung und die Datensicherung pro Übertragung muss sowohl die Bandbreite der NAK-Retransmissions zwischen Sender und Empfängern als auch die zwischen Repetitoren und Empfängern einbezogen werden. Dies erfolgt durch die Konfiguration eines festen erlaubten Prozentsatzes der Gesamtbandbreite für jeden der erwähnten Datensicherungsmechanismen pro Channel. Dabei wird der Prozentsatz der Bandbreite für FEC und Packet Retransmission von der im Sender konfigurierten maximalen zulässigen Bandbreite für den Channel abgezogen, wobei die FEC-Bandbreite konstant von FEC genutzt wird, während die für Packet Retransmission nur im Bedarfsfall abgezogen wird, d. h. die Bandbreite steht sonst für die eigentliche Übertragung zur Verfügung.

Die Bandbreite für Repetitor Packet Retransmission wird im Gegensatz dazu auf die momentane Bandbreite des Channels aufgerechnet. Der Anteil an Repetitor Packet Retransmission kann in allen Repetitorgruppen individuell konfiguriert werden.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

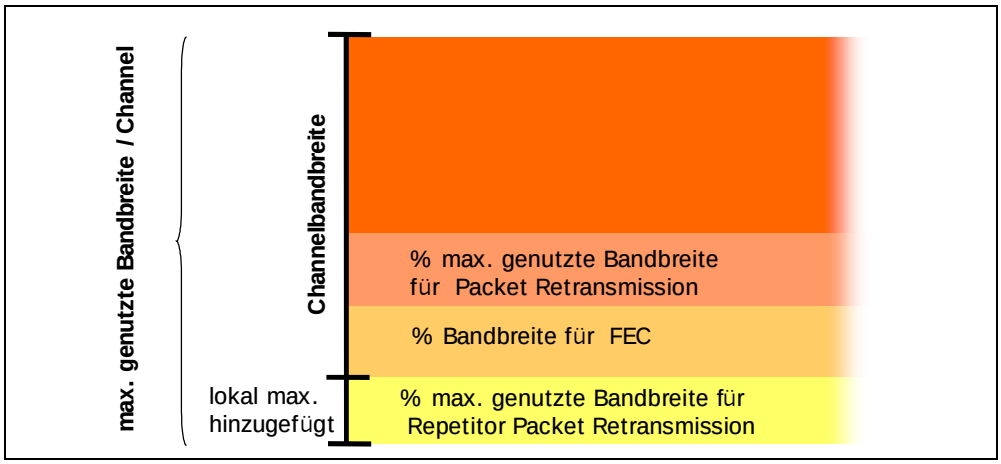

Abbildung 11: Einteilung der Channelbandbreite unter Berücksichtigung von Datensicherungsmechanismen

6.4.3 Steuerung von Retransmissions

tq®-TELLICAST kann so konfiguriert werden, dass die Empfänger den vollständigen Empfang der übertragenen Dateien durch Acknowledgements bestätigen. Fehlen die Acknowledgements von einzelnen oder mehreren Empfängern, kann die Übertragung von tq®-TELLICAST wiederholt werden. Dabei kann in den Job Files eine maximale Anzahl an Übertragungswiederholungen konfiguriert werden.

Für die Multicastübertragung in MDiS ist es wünschenswert, dass die Steuerung der Retransmissions von MD und nicht von tq®-TELLICAST vorgenommen wird. tq®-TELLICAST überträgt daher die Daten nur einmal. Die eingegangenen Acknowledgements für diese Übertragung werden von MD aus dem Job File ausgelesen, sobald dieses im „jobs/done“-Verzeichnis erscheint. Retransmission an die Empfänger, die den Empfang der Daten nicht bestätigt haben, erfolgt durch die Generierung eines neuen Job Files durch MD. Dies hat die folgenden Vorteile gegenüber der automatischen Retransmission durch tq®-TELLICAST innerhalb eines einzigen Jobs:

- Der Status der Übertragung wird nach jeder einzelnen Übertragung an MD weitergegeben und kann an die GUI nach Anfrage weitergeleitet werden.
- Die Priorität der Retransmissions kann gegenüber der Priorität der Erstübertragung herabgesetzt werden. Da für den Wetterdienst besonders die letzten, aktuellen Daten wichtig sind, können dadurch neu eingehende Sendeaufträge vor den Retransmissionen alter Daten übertragen werden, weil sie die höhere Priorität besitzen.

Die Empfängerliste für die Retransmission kann gezielter auf die Empfänger beschränkt werden, die keine Acknowledgements gesendet haben.

6.4.4 Realisierung des ISDN-Backups

Die Daten werden über 2 verschiedene Multicastadressen auf 2 Channeln von tq®-TELLICAST gesendet:

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

- Die wichtigsten Daten werden konstant über einen Multicast-Channel gesendet, der sowohl über ATM als auch über ISDN geroutet werden kann, und der im normalen Betrieb über ATM und bei Bedarf automatisch über ISDN geroutet wird.
- Die weiteren Daten werden über einen zweiten Channel übertragen, der nur über ATM geroutet wird, daher über ISDN/Ethernet nicht empfangen werden kann und somit auch keine Bandbreite der ISDN-Verbindung in Anspruch nimmt.

Solange über ATM empfangen werden kann, gehen alle Daten über ATM ohne Verzögerung bei den Empfängern, d. h. den Daten-Servern, ein. Wenn ATM ausfällt, ist über das ISDN-Backup eine Notversorgung mit den wichtigsten Daten gewährleistet.

Die Multicastempfänger auf den Daten-Servern, die im Normalbetrieb auf den Empfang über das ATM-Interface eingestellt sind, können den Ausfall der ATM-Übertragung über den Announcementchannel von tq®-TELLICAST registrieren. Der Announcementchannel wird immer von allen Empfängern konstant empfangen. Daher kann er als Indikator für den Ausfall des Netzes dienen. Registriert der Empfänger, das er den Announcementchannel über ATM nicht mehr empfängt, so wird der Empfang über das Ethernet-Interface auf ISDN umgestellt.

6.5 BANDBREITENMANAGEMENT BEI MEHREREN SENDERN

Für alle Sender des Systems wird ein gemeinsames Bandbreitenmanagement realisiert. Ein Sender wird dabei zum Master-Sender und reguliert die Gesamtbandbreite. Die anderen Sender teilen dem Master-Sender ihren Bandbreitenbedarf mit und bekommen von diesem entsprechend der aktuellen Verfügbarkeit Bandbreite zugewiesen.

Welcher Sender das Bandbreitenmanagement übernimmt, wird zwischen den Sendern automatisch ausgehandelt. Jeder Sender bekommt über die Konfigurationsdatei eine Priorität zugewiesen. Die Sender tauschen sich über die Priorität aus und weisen die Master-Funktion dem Sender mit der höchsten Priorität zu.

Der Master-Sender teilt den anderen Sendern über den Announcementchannel in regelmäßigen Abständen mit, an welche Adresse die Anfragen für Bandbreite gesendet werden sollen. Wenn dieses Signal ausbleibt, wird automatisch ein neuer Sender als Master-Sender festgelegt.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.6 SCHNITTSTELLE ZUM MDIS-GUI

Die Schnittstelle zwischen dem Sender-seitigen MD / tq®-TELLICAST und der Sendekontrolle über das GUI (Graphic User Interface von AFD) bildet ebenfalls das MD.

Die GUI (oder alternativ auch ein anderer Prozess) verbinden sich über einen TCP-Server-Port mit dem MD und rufen die Daten über ein ASCII-basiertes Protokoll ab. MD lässt dabei eine Abfrage nur von Rechnern zu, die in einer GUI-Rechnerliste des MD eingetragen sind.

MD liest die an das GUI zu übermittelnden Daten über den Sendeerfolg der Übertragung aus den tq®-TELLICAST Job Files im „jobs/done“-Verzeichnis aus.

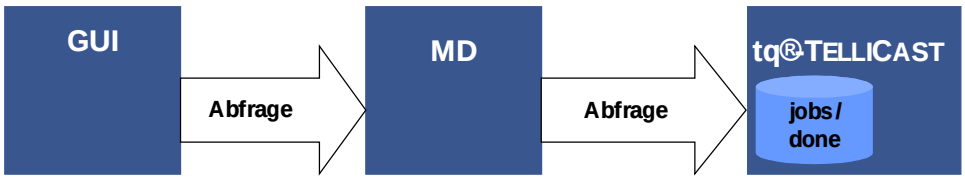

Abbildung 12: Abruf von Informationen über die Multicast-Übertragung durch das GUI

Ein zentrales GUI kann die notwendigen Daten der einzelnen MD-Prozesse entweder ebenfalls direkt oder (performanter) analog zu AFD von den einzelnen GUIs abfragen.

Von MD werden die Informationen über den Sendeerfolg wie folgt übermittelt:

- pro Empfänger
- als noch zu übertragende Dateien
- als noch zu übertragende Bytes

Unabhängig vom Rhythmus und Art der vom GUI angefragten Informationsübergabe überträgt MD bei jeder Rekonfiguration eine aktuelle Empfängerliste und Channel-Liste an das GUI.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

6.7 REALISIERUNG IM RAHMEN DES PROJEKTES

Nach der Konzeptionierung erfolgte Mitte des Jahres 2001 die Bereitstellung des um den Prozess „Multicast Distribution Layer“ erweiterten Multicastübertragungssystems tq®-TELLICAST durch die Firma Tellique. Aufgabe des DWD war anschließend der Test dieser Software und darauf folgend die Aufnahme der routinemäßigen Datenverteilung über Multicast.

6.7.1 Tests

Die Funktionsfähigkeit der gelieferten Software wurde durch die Firma Tellique auf zwei O200-Rechnern – einem Sender und einem Empfänger – vorgeführt. Nach den ersten Konfigurationsversuchen und Tests wurde die Testumgebung um drei Linux-PCs erweitert. Die fünf Rechner - eine O200 dient als Sender - befinden sich in drei Subnetzen und bieten so eine Testplattform für die betriebliche Konfiguration des DWD. Bei den einführenden Testläufen ermittelte Fehler wurden von Tellique nach Meldung beseitigt.

Die Tests auf Datenintegrität der Multicast-Verteilung sind noch nicht beendet. Hierfür wurde ein Testsystem entwickelt, dass eine definierte Anzahl von binären Dateien generiert und diese dem MDiS übergibt. Auf den Empfängern wird nach der Datenübertragung die Vollständigkeit und Integrität der Dateien automatisch geprüft. Dieser Test wird mehrfach und mit unterschiedlichen Dateigrößen durchgeführt. Auch wechselt der Sender zwischen den fünf Rechnern.

Ein Bestandteil der Testprozeduren sind Durchsatztests, in denen die Leistung des Multicastsystem gemessen und teilweise Vergleichsmessungen mit der DWD-eigenen operationellen Fileverteilung per FTP-Protokoll vorgenommen werden.

In einer zweiten Testgruppe werden Störungen im Netz und auf den Rechnern simuliert und die korrekte Reaktion der Multicastdatenverteilung auf diese Problemsituationen verifiziert. Diese Test sind für den DWD besonders wichtig, da im Gegensatz zu üblichen Anwendungen von Multicast-Verteilung (z.B. Video-Konferenzen), bei denen der Verlust einzelner UDP-Pakete akzeptiert werden kann, bei der Dateiverteilung im DWD mittels Multicast keine Datei verloren gehen bzw. verfälscht werden darf. Deshalb wird dieser Test mit aller Sorgfalt durchgeführt und ist noch nicht abgeschlossen.

6.7.2 Festlegung der Multicast-Channel

Die Festlegung der Multicast-Channel ist ein iterativer Prozess, bei dem die meteorologisch-technischen Anforderungen (z.B. Zeitverhalten der Datenverteilung; Prioritäten bestimmter Datenarten) und die heterogene Netzwerkleistungssituation (unterschiedlich leistungsfähige Anbindung von Dienststellen, die teilweise dieselben Daten benötigen) abgestimmt werden müssen. Für die für das Projekt vorgesehene Versorgung der Regionalzentralen mittels Multicast scheint die Festlegung zunächst eine einfache Aufgabe zu sein. Für eine Planung des Gesamtnetzes ist aber eine differenzierte Planung der Multicastkanäle erforderlich, bei der das Auslastungsverhalten der nur mit 64 kBit/sec realisierten Stichverbindungen mit zu berücksichtigen ist. Da die bisherige Fileverteilung die Daten entspre-

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

chend ihrer jeweiligen Priorität über eine oder nur maximal einige wenige logische Verbindungen von Knoten zu Knoten transportiert, fehlen Erfahrungswerte über das Stauverhalten der notwendigen größeren Anzahl von Multicastdatenströmen an den Übergangspunkten zwischen schnellen und langsamen Netzabschnitten. Die Festlegung der Multicastchannel wird zur Zeit als Vorbereitung für die Inbetriebnahme parallel zu den Tests vorgenommen und ist bisher noch nicht abschließend erfolgt.

6.8 INBETRIEBNAHME

Da die Testphase noch nicht abgeschlossen werden konnte, kam eine Inbetriebnahme während der Laufzeit des Projektes noch nicht in Frage. Noch nicht entschieden ist außerdem die Frage, ob die Versorgung der kleineren Standorte von der Zentrale in Offenbach aus erfolgen soll, oder ob es besser ist, in jeder Regionalzentrale einen Multicastsender einzurichten, der die an dieser Regionalzentrale angeschlossenen Standorte des Sekundärnetzes versorgt. Im ersten Fall würden die Daten, um den teilweise engen zeitlichen Versorgungsanforderungen gerecht zu werden, zweimal gesendet werden – einmal für die Regionalzentralen und einmal mit geringerer Bandbreite für die Standorte des Sekundärnetzes. Im zweiten Fall würde ein deutlich höherer Konfigurationsaufwand – für jede Regionalzentrale ein eigener Channel zur Versorgung der angeschlossenen Standorte – und eine zusätzliche Übergabefunktion an den Multicastsender an der Regionalzentrale benötigt werden.

6.9 BEWERTUNG DES ERGEBNISSES

Die bisher vorliegenden Testergebnisse zeigen, dass das unter MDiS entwickelte Multicastsystem für eine Datendistribution innerhalb des DWDs geeignet ist. Allerdings erfordert eine erfolgreiche Nutzung noch einen Konfigurations- und Planungsaufwand für die Umstellung des Systems von Seiten des DWD.

Ausgangspunkt der Planungen des DWD für den Betriebseinsatz waren die effektivere Nutzung der Netzwerk- und Rechnerresourcen und die Einsparung von Konfigurations- und Betreuungsaufwand durch eine Zentralisierung der Steuerung der Datenverteilung.

Die als primäres Projektziel geplante Umstellung der Datenverteilung von der Zentrale zu den Servern auf den Dienststellen bringt dem DWD zunächst nicht den erwarteten Nutzen, da der Aufwand für die Datenverteilung auf die unterschiedlichen Anwendungen auf den Servern bzw. die Unterverteilung auf die Rechner der Dienststellen bleibt. Die durch die Aufgabe des Hub-To-Hub Systems eingesparten Betreuungs-aufwände werden voraussichtlich durch den Betreuungsaufwand für die Multicast-Anwendung kompensiert werden. Eine wirksame Minderung des Aufwandes kann nur durch die direkte Multicastversorgung der Anwendungen mit den für sie relevanten Informationen erreicht werden. Die dafür erforderliche Migration der DWD-Anwendungen war im Projekt nicht vorgesehen und ist nur auf mittlere Sicht umsetzbar. Zur Zeit wird eine Zwischenlösung realisiert, bei der Multicastempfänger auf allen zu versorgenden Rechnern installiert und lokal die Zuordnung der Daten zu den Anwendungen vorgenommen wird.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

Eine Minderung der Netzlast im Primärring gegenüber der Unicast-Datenverteilung war von vorneherein nicht zu erwarten, da das Hub-to-Hub-System des DWD eine Art „simulierte Multicastverteilung“ im Primärring bewirkte. Die Daten wurden jeweils von Niederlassung zu Niederlassung weitergereicht, anstatt von der Zentrale an jede Niederlassung einzeln versendet zu werden.

Seit dem Sommer 2001 wurden in einer Studie des BMVBW die Möglichkeiten zur Harmonisierung der europäischen meteorologischen Satellitenverteildienste untersucht. Dabei stellte sich heraus, dass die Überlagerung der terrestrischen Netze des DWD durch eine Multicast-Datenverteilung über eine hochverfügbare Satellitenstrecke für die Grundversorgung der DWD-Dienststellen wirtschaftlich vorteilhaft ist. In Konsequenz beauftragte der DWD-Vorstand der Aufbau eines derartigen Systems bis Ende 2002. Für dieses Projekt wird die mit MDiS geschaffene Datenverteilung so modifiziert, dass die Verteilanforderungen des DWD durch einen Verbund von terrestrischen und satellitengestützten Multicastverbindungen abgedeckt werden können. Durch die Verwendung von MDiS sowohl im terrestrischen Netz als auch auf der Satellitenstrecke kann eine optimale Kompatibilität beider Datenverteilsysteme erreicht werden, was den Verwaltungsaufwand erheblich senken kann.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

7 PROJEKTERGEBNIS

Ziel des Projektes MDiS ist der Piloteinsatz eines gesicherten Multicast-Dateiverteilsystems auf Basis des Transportprotokolles MTP/SO.

Pilotnutzer im Projekt ist der DWD. Der DWD bietet sich als Pilotnutzer an, da über das AFD-System in den Standleitungen des Primärnetzes und des Sekundärnetzes des DWD große Mengen komplexer Daten zwischen den Niederlassungen ausgetauscht werden, wobei an mehrere Standorte identische Daten übertragen werden.

Die Multicast-Dateiverteilung für den DWD wurde mit tq®-TELLICAST FileBroadcast realisiert, einer von Tellique auf Basis MTP/SO entwickelten Software für gesicherte Datenverteilung über Multicast. tq®-TELLICAST wurde über einen Adaptationslayer (Multicast-Distribution, MD) in die übliche Datenverteilung des DWD über AFD integriert.

In ersten Tests konnte während der Laufzeit des Projektes die Funktionstüchtigkeit des Systems nachgewiesen werden. Der DWD hat bisher jedoch die Multicast-Datenverteilung noch nicht in den laufenden Betrieb integriert. Da die Anforderungen an die Sicherheit und die zeitkritische Übertragung beim DWD besonders hoch sind, legt der DWD Wert darauf, zunächst langfristigere Tests durchzuführen. Zudem erfordert die Etablierung des neue System zunächst erhöhten Konfigurationsaufwand in der Umstellungsphase, um eine optimale Anpassung sämtlicher Parameter zu gewährleisten. Der DWD wird deshalb das System schrittweise nach den Langzeittests einführen. Die bisher vorliegenden Testergebnisse sind jedoch schon so positiv, dass der DWD beschlossen hat, die unter MDiS verwendete Software zusätzlich zur Datenverteilung über die terrestrischen Netze auch für die Satellitendistribution einzusetzen. Damit wäre eine optimale Kompatibilität zwischen den Systemen gegeben.

Für die Integration der Multicast-Datendistribution in das AFD-System des DWD wurde die von Tellique entwickelte Software tq®-TELLICAST verwendet. Um den universellen Einsatz von MTP/SO auch außerhalb der Einbindung in tq®-TELLICAST zu demonstrieren, wurde auf Grundlage des weit verbreiteten Unicast-Dateiverteilsystems *rdist* ein Tool zur Multicast-Dateiverteilung, *mdist*, entwickelt. Durch die Entwicklung von *mdist* ist ein Tool entstanden, das es Anwendern, die *rdist* nutzen, ermöglicht, Daten über Multicast zu übertragen, ohne die schon vorhandene Konfiguration von *rdist* ändern zu müssen. Dadurch kann gesicherte Multicastübertragung für eine Vielzahl von Nutzern des Wissenschaftsnetzes des DFN interessant werden.

Weitere Anwendungsmöglichkeiten, die sich für Nutzer des Wissenschaftsnetzes ergeben, sind unter anderem die Distribution von Software sowie die Realisierung von FTP-Mirror und Web-Mirror. Neben der Integration in Anwendungen bietet sich IP-Multicast auch als Unterstützung der zentralen DFN-Infrastruktur, z. B. zur Verteilung von NetNews und zum Web-Caching an. Um die Nutzung von Multicast im Bereich des B-WIN zu fördern, wird das Protokoll MTP/SO auf Anforderung den DFN-Mitgliedern zur Verfügung gestellt, wenn Projekte unter Verwendung von Multicast innerhalb des B-WIN realisiert werden sollen. Über eine BSD-Socket-

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

ähnliche Schnittstelle können so verschiedenartige Anwendungen von den Vorteilen von IP-Multicast profitieren.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

8 ABKÜRZUNGSVERZEICHNIS/GLOSSAR

A

accounting.dat	Datei von tq®-TELLICAST, in die das System während der Übertragungen die Informationen schreibt, welche Daten an wen übertragen wurden.
ACK	Acknowledgement, Bestätigung über den Datenempfang, die der Empfänger an den Sender schickt.
AFD	Automated File Distributor; Dateiübertragungssystem des DWD.
Announcement-Channel	Logischer Kanal (Channel), über den tq®-TELLICAST Übertragungsaufforderungen und Datenschlüssel an die Empfänger verteilt.
ASCII	American Standard Code for Information Interchange.
ATM	Asynchronous Tranfer Mode; Zellbasiertes Übertragungsverfahren.

B

bit/s	Bit pro Sekunde
B-WiN	Breitband-Wissenschaftnetz des DFN-Vereins.
bzw.	Beziehungsweise

C

Channel	Logischer Kanal, über den tq®-TELLICAST Daten versendet. tq®-TELLICAST teilt physikalische Kanäle in mehrere Channel auf. (Auf IP-Ebene gleichzusetzen mit einer Multicast-Gruppe.)

D

d. h.	das heißt
DFN-Verein	Verein zur Förderung des Deutschen Forschungsnetzes, Betreiber des WiN und B-WiN.
DWD	Deutscher Wetterdienst

E

etc.	Etcetera; und so weiter
Ethernet	LAN-Protokoll IEEE 802.3

F

FEC	Forward Error Correction; Korrektur von Übertragungsfehlern ohne Verwendung eines Rückkanals durch Berechnung von fehlenden Datenpaketen über Redundanzpakete.
FTP	File Transfer Protocol. Eine Standardanwendung für TCP/IP, die nur die Fileübertragung und keinen Filezugriff beinhaltet.

G

GByte	Gigabyte
ggf.	Gegebenenfalls

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

	GUI	Graphical User Interface; hier: Kontrolloberfläche des AFD-Systems des DWD.
I		
	IETF	Internet Engineering Task Force – Standardisierungsorganisation des Internets.
	Invitation	Empfangsaufforderung für eine Datenübertragung auf einem bestimmten Channel, die der tq®-TELLICAST Sender über den Announcementchannel an die tq®-TELLICAST Empfänger sendet.
	IP	Internet Protokoll
	IRTF	Arbeitsgruppe der IETF zur Entwicklung von Standards zur gesicherten Datenübertragung.
	ISDN	Integrated Services Digital Network; digitales Datenübertragungsnetz.
K		
	kbit/s	Kilobit pro Sekunde
L		
	LAN	Local Area Network
	log.dat	Datei von tq®-TELLICAST, in die Fehlermeldungen und Systemstatusmeldungen während des Betriebs von tq®-TELLICAST geschrieben werden.
M		
	MByte/s	Megabytes pro Sekunde
	MDiS	Multicast Distribution System; System zur gesicherten Übertragung von Multicast, das im Rahmen eines DFN-Projektes vom DWD, der Tellique GmbH und dem Technologiezentrum Informatik der Universität Bremen beim Deutschen Wetterdienst in Betrieb genommen wird.
	mdist	Aus *rdist* und MTP/SO im rahmen von MDiS entwickeltes Tool zur Multicast-Dateiverteilung.
	MTP/SO	Multicast Tranport Protocol / Self Organizing; von der Tellique GmbH vertriebenes Transportprotokoll zur sicheren Datenübertragung über Multicast auf der Basis des RFC 1301.
	Multicast	Mehrpunktübertragung. Übertragung von einem Punkt zu einer Gruppe.
	Multicast-Backbone	Multicastfähiger „Backbone“ im Internet.

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

N

NAK	Negative Acknowledgement. Übertragungs-kontrolle durch Rückmeldung vom Empfänger, wenn die erwarteten Datenpakete vom Sender nicht empfangen wurden.

P

Primärnetz	Standleitungen mit einer Bandbreite von 34 KBit/s, die die Zentrale des DWD in Offenbach und die 6 Niederlassungen mit Regionalzentrale ringförmig miteinander verbinden.

R

rdist	Programm zur Unicast-Dateiverteilung unter UNIX.

S

Sekundärnetz	Standleitungen geringerer Bandbreite, die kleinere Standorte des DWD mit den Niederlassungen mit Regionalzentralen verbinden.
SMTP	Simple Mail Transfer Protokoll

T

TCP	Transmission Control Protocol. Verbindungsorientiertes Transportprotokoll, das in Verbindung mit IP eingesetzt wird.
tq®-TELLICAST	Multicastübertragungs-Software der Tellique Kommunikationstechnik GmbH.
TTL	time to live; Anzahl der Router, die das Datenpaket weiterleiten. Jeder Router zählt die TTL des Paketes um 1 herunter. Ist die TTL Null, wird das Paket nicht mehr weitergeleitet.

U

u. a.	unter anderem
UDP	User Datagram Protocol. Verbindungsloses Transportprotokoll, das auf IP aufsetzt. UDP bietet eine ungesicherte Übertragung von Datagrammen.
Unicast	Punkt zu Punkt Datenübertragung.

V

VC	Virtual Channel

W

WiN	Wissenschafts-Netz des DFN-Vereins.

Z

z. B.	zum Beispiel

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de

9 ABBILDUNGSVERZEICHNIS

PDF wurde mit FinePrint pdfFactory-Prüfversion erstellt. http://www.context-gmbh.de