

Netze und Protokolle für das Internet



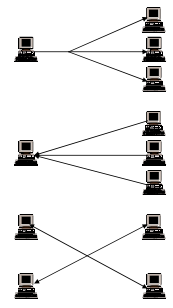
10. Transportprotokolle zur Gruppenkommunikation

Inhalt

- Kommunikationsformen
- Zuverlässigkeitsklassen
- Anwendungen von zuverlässigem Multicast
- TCP-Eigenschaften
- Multicast Transport Protocol
- Reliable Multicast Protocol
- Scalable Reliable Multicast
- Baum-basierte Ansätze
 - Reliable Multicast Transport Protocol
 - Reliable Multicast proXies
 - MTCP
 - Active Reliable Multicast
- Vorwärtsfehlerbehebung
- Staukontrolle
- IETF WG Reliable Multicast Transport

Kommunikationsformen

- 1:1 Unicast
- 1:n Multicast
- n:1 Concast
- n:m Multipeer



Zuverlässigkeitsklassen

- unzuverlässig
 - keine Auslieferungsgaranten
 - ggf. Übertragung redundanter Information
 - halbzuverlässig
 - statistisch zuverlässig
 - Schwellwert gibt an, wieviele Gruppenmitglieder die Daten innerhalb eines bestimmten Zeitintervalls erhalten müssen.
 - Gruppengröße muss bekannt sein.
 - k-zuverlässig
 - garantiert den Empfang der Daten an k Gruppenmitglieder ($k \leq n$, n = Gruppengröße)
 - zuverlässig
 - Daten werden fehlerfrei, ohne Duplikate und in der korrekten Reihenfolge bei allen Gruppenmitgliedern ausgeliefert.
- Nach Ablauf eines Zeitintervalls können Korrekturmaßnahmen ergriffen werden, z.B. Übertragungswiederholung

Anwendungen von zuverlässigem Multicast

- Push Technologien
- Software Aktualisierungen
- Cache Aktualisierungen
- Verteiltes Rechnen
- Computer Supported Cooperative Work (CSCW)
- Application Sharing

Shared Whiteboard



Zuverlässige Multicast Kommunikation im Internet

- TCP ist nur für 1:1-Verbindungen geeignet
- Multicast-Unterstützung im Internet durch UDP
- UDP ist unzuverlässig und besitzt keine Stau- bzw. Flusskontrolle.
- Transportprotokolle im Internet sollten TCP-Eigenschaften aufweisen, besonders hinsichtlich Staukontrolle → TCP-Freundlichkeit
- Multicast-Transportprotokolle setzen meist auf UDP auf und sind in Anwendungen integriert.

Anwendung
zuverlässiges Multicast-Protokoll
UDP
IP
Netz

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

7

TCP-Eigenschaften

- Flusskontrolle mit Fenstermechanismus
- Staukontrolle mit Slow Start Mechanismus
- Fehlerkontrolle
 - Folgenummern
 - Prüfsumme
 - Quittierungsnummern
 - Übertragungswiederholung
- Reihenfolgetreue

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

8

Klassifikation zuverlässiger Multicast-Transportprotokolle

- Sender-initiierte Protokolle
 - Quelle verwaltet Informationen über alle Empfänger
 - Empfänger senden ACKs und NACKs
 - Probleme: Sender muss Empfänger kennen, ACK-Verarbeitung
 - Beispiele: XTP, NETBLT
- Empfänger-initiierte Protokolle
 - Verantwortung für Zuverlässigkeit beim Empfänger
 - Anforderung von Übertragungswiederholungen durch NACKs
 - Beispiel: SRM
- Baum-basierte Protokolle
 - Unterteilung der Empfänger in Teilgruppen, Baumstrukturen
 - Beispiel: RMTP
- Ring-basierte Protokolle
 - ausgezeichnete Station (Token-Halter) zur Generierung von ACKs, Token-Weitergabe
 - Beispiel: RMP

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

9

Multicast Transport Protocol

- halbzuverlässiger Multipeer-Dienst mit Ordnungserhaltung
- Rollen der Gruppenmitglieder
 - Master
 - entscheidet über Aufnahme neuer Mitglieder
 - vergibt Senderecht
 - überwacht zuverlässige Datenübertragung und Ordnungserhaltung
 - kann auch als Produzent wirken
 - Produzent (Sender + Empfänger)
 - Konsument (Empfänger)
- Kommunikationsgruppe = Web
 - Aufbau des Webs: Senden von JOIN-REQUEST an Multicast-Gruppe
 - Master antwortet per Unicast mit JOIN-CONFIRM oder JOIN-DENY
 - Bei ausbleibender Antwort kann ein Teilnehmer die Master-Rolle übernehmen.
 - Freiwilliges oder erzwungenes Verlassen des Web
 - Master löst Web mit QUIT-REQUEST auf.

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

10

Vergabe der Senderechte bei MTP

- Produzent darf Daten nur dann versenden, wenn er im Besitz des Senderechts ist.
- Produzent muss beim Master per Unicast Senderecht anfordern.
- Master erteilt Senderecht (Token) unter Angabe einer Nachrichten Sequenznummer.
- Pakete einer Nachricht werden mit Paket Sequenznummer gekennzeichnet.
- Produzent gibt Token explizit durch Setzen von End of Message flag in letzter Nachricht frei.

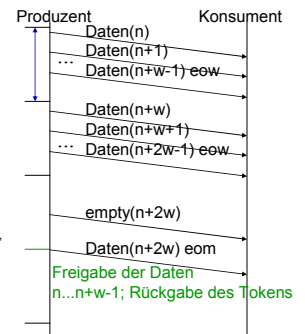
SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

11

MTP-Datentransfer

- Vereinbaren von Parametern beim Aufbau eines Webs
 - Heartbeat
 - mindestens ein Empty-Paket pro Zeitintervall
 - Window
 - Anzahl der Pakete, die in einem Heartbeat-Intervall gesendet werden dürfen
 - Beispiel: $w=3$
 - Retention
 - Anzahl der Heartbeat-Intervalle, während der ein Produzent gesendete Nutzdaten für Übertragungswiederholungen bereit halten muss
 - Beispiel: $r=2$



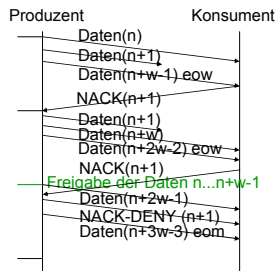
SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

12

MTP-Fehlerkontrolle

- selektive Übertragungswiederholung
- Retention-Wert stellt Zuverlässigkeitsmass dar.
- Beispiel: $r=1$
- Übertragungswiederholung (per Multicast) unterliegen der Flusskontrolle (Ratenkontrolle)
- Problem: Quittungsimplosion



SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

13

MTP: Ordnungs- und Konsistenzhaltung

- Nachrichten-Sequenznummer legt Auslieferungsreihenfolge fest.
- Gruppenmitglieder sollten Daten einheitlich ausliefern → Konsistenz
- Master entscheidet, ob eine Nachricht gültig ist oder nicht, d.h. nur beim Master vollständig eingetroffene Nachrichten werden als gültig erklärt.
- Problem: Master erhält Nachricht, Empfänger wegen zu kleinem Retention-Intervall aber nicht.
- Zustände der Nachrichten beim Master
 - akzeptiert
 - ausstehend
 - zurückgewiesen
- Statusvektor (Teil jeder Nachricht) enthält Zustand der letzten 12 Nachrichten.
- Neue Nachricht darf nur gesendet werden, wenn älteste Nachricht nicht mehr den Zustand „ausstehend“ hat.

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

14

Reliable Multicast Protocol

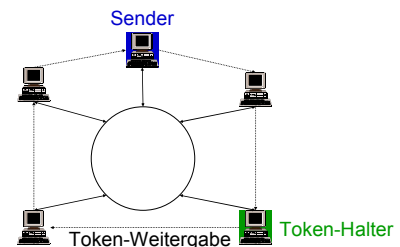
- zuverlässige und ordnungserhaltende n:m-Kommunikation
- Anordnung von Gruppenmitgliedern in Ring
- Senden durch Nicht-Gruppenmitglieder über Proxy-Mitglied
- Senden von positiven / negativen Quittungen (ACK/NACK) per Multicast
- Ausgezeichnetes Mitglied (Token-Halter) hat Aufgabe, Daten von mehreren Sendern zu serialisieren und positive Quittungen zu erzeugen.
- Token rotiert zwischen Ring-Mitgliedern (→ Fehlertoleranz)
- dynamische Aufnahme und Löschen von Ring-Mitgliedern
- verschiedene Auslieferungsdienste

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

15

RMP-Szenario



SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

16

RMP-Dateneinheiten und Quittungen

- Dateneinheiten enthalten Tupel
 - Sender-ID
 - Sequenznummer des Senders
 - gewünschter Auslieferungsdienst
- Quittungen enthalten
 - globale Sequenznummer (Zeitstempel)
 - Token-Halter ordnet jeder quitierten Dateneinheit eine globale Sequenznummer zu.
 - Tupel der quitierten Dateneinheiten
 - Adresse des bisherigen und nächsten Token-Halters

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

17

RMP-Protokoloperationen

- Senden
- positive Quittierung (ACK)
 - Multicast durch Token-Halter
 - kann mehrere Pakete verschiedener Sender quittieren
 - enthält Zeitstempel
- negative Quittierung (NACK)
 - Mitglied mit fehlenden Daten sendet NACK per Multicast
- Übertragungswiederholung
 - durch Token-Halter oder anderen Knoten
- Auslieferung
 - abhängig von Auslieferungsdienst
- Token-Weitergabe
 - mit Senden eines ACK
 - Neuer Token-Halter muss alle Nachrichten erhalten haben.
- Wechseln der Mitgliedschaft
 - List Change Request
 - Token-Halter erzeugt neue Liste und bestätigt Aufnahme

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

18

RMP-Auslieferungsdienste

- unzuverlässig
 - einmalige, mehrmalige oder keine Auslieferung, ungeordnet
- zuverlässig
 - mindestens eine Auslieferung, aber nicht geordnet
- Quellen-geordnet
 - mindestens 1 Auslieferung, in der gleichen Reihenfolge wie von einem Sender erzeugt, keine Ordnung zwischen den Sendern
- Total-geordnet
 - Auslieferungsreihenfolge von verschiedenen Sendern ist bei allen Empfängern identisch.
- k-elastisch
 - totale Ordnung, Auslieferung bei mindestens k Empfängern
- Mehrheits-elastisch
 - K-elastisch mit $k > N+1/2$, N: Anzahl der Empfänger
- Total-elastisch
 - K-elastisch mit $k = N$

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

19

RMP-Auslieferungsoperationen

- ungeordnet
 - sofortige Auslieferung nach Erhalt einer Nachricht
- Quellen-geordnet
 - Pakete einer Quelle werden ausgeliefert sobald alle Pakete mit niedrigerer Sequenznummer empfangen wurden.
- Total-geordnet
 - Pakete werden ausgeliefert, wenn alle Pakete mit kleinerem Zeitstempel ausgeliefert wurden.
- k-elastisch
 - Pakete mit kleinerem Zeitstempel wurden ausgeliefert und Token wurde k-mal weitergegeben
- Mehrheits-elastisch
 - Pakete mit kleinerem Zeitstempel wurden ausgeliefert und Token wurde $N/2$ -mal weitergegeben
- Total-elastisch
 - Pakete mit kleinerem Zeitstempel wurden ausgeliefert und Token wanderte einmal um den Ring

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

20

Scalable Reliable Multicast

- Menge von in Anwendung zu integrierende Funktionen, z.B. shared whiteboard
- Komponenten
 - anwendungsabhängig: eindeutige Bezeichnung der Dateneinheiten
 - anwendungsunabhängig: Kontrollalgorithmen
- Anwendung muss Ordnung selbst herstellen.
- Empfänger sind für die zuverlässige Zustellung selbst verantwortlich.
- Ratenbasierte Flusskontrolle durch Sender
- Übertragungswiederholungen sollten durch nächsten benachbarten Empfänger erfolgen.
- Protokollnachrichten
 - Repair Request
 - Repair (Übertragungswiederholung)

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

21

SRM: Fehlererkennung und -behebung

- Empfänger erkennen nicht erhaltene Pakete anhand von Lücken im empfangenen Sequenznummernbereich sowie durch periodisch gesendete Statusnachrichten (inkl. höchste empfangene Sequenznummer und Zeitstempel) der anderen Gruppenmitglieder
- Repair Requests werden nach Timeout $[c_1 \cdot d_R, (c_1+c_2) \cdot d_R]$ gesendet, d_R = geschätzte Einwegverzögerung zwischen Sender und Empfänger R
- Beim Empfang einer Repair-Request-Nachricht von Empfänger X wählt Empfänger Y einen Timeout $[c_3 \cdot d_{XY}, (c_3+c_4) \cdot d_{XY}]$, d_{XY} = geschätzte Einwegverzögerung zwischen X und Y
- Bei Ablauf des Timeout ohne Empfang einer Repair-Nachricht wird Repair per Multicast gesendet.
- Ziel: 1 Repair, 1 Repair-Request

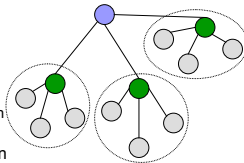
SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

22

Reliable Multicast Transport Protocol

- Protokoll für einen Sender
- Aufbau einer Hierarchie zum
 - Reduzieren von ACK/NACK-Nachrichten
 - Reduzieren von Verzögerungen durch lokale Übertragungswiederholungen
- Empfänger werden in lokale Regionen gruppiert.
- Jede Region besitzt ausgezeichneten Empfänger (designated receiver, DR), welcher die lokale Region repräsentiert.
- mehrere Hierarchieebenen möglich
- Sender und DRs senden Kontrollnachrichten mit gleichen TTL-Werten
→ Auswahl der DRs mit grösstem TTL-Wert



SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

23

RMTP-Protokolloperationen

- Empfänger senden periodisch Statusnachrichten zum DR
- DR sendet Statusnachrichten (Kombinationen von ACK / NACK) zum Sender
- Statusnachrichten enthalten Sequenznummer des ersten nicht erhaltenen Pakets + Bitvektor über Status der folgenden Pakete
- Rate der Statusnachrichten hängt von RTT zwischen Empfänger und DR bzw. zwischen DR und Sender ab.
- Lokale Übertragungswiederholungen nach Timeout per Unicast / Multicast abhängig von Empfänger-Anzahl
- Raten-/Fenster-basierte Flusskontrolle
- Empfänger bzw. DR kann auf bestimmte Daten verzichten, z.B. nach Beitritt oder Auflösung einer Netzpartitionierung

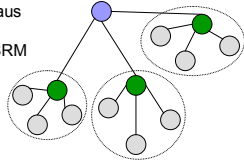
SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

24

Weitere Baum-basierte Ansätze

- Reliable Multicast proXies
 - RMXs bilden Spanning Tree (vgl. Bridges)
 - RMXs tauschen Daten über TCP aus → Staukontrolle
 - Lokale Multicast-Verteilung über SRM
- MTCP
 - Zwischenknoten (Sender Agents) melden Staukontrollinformationen stromaufwärts
 - Congestion Window (cwnd): Wert basierend auf Schätzung der minimale Bandbreite
 - Anzahl noch nicht quittierter Daten in stromabwärts liegenden Knoten (twnd)
 - Aggregation: min (cwnd), max (twnd)



SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

25

Active Reliable Multicast

- Ansätze
 - Active Reliable Multicast
 - Reliable Multicast Active Network Protocol
- Router entlang des Multicast Bums
 - speichern Daten im lokalen Cache
 - aggregieren ACKs
 - unterdrücken NACKs
 - führen lokale Übertragungswiederholungen aus.
- Code: in band oder out of band

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

26

Vorwärtsfehlerbehebung

- Die meisten Multicast-Transportprotokolle basieren auf Übertragungswiederholung (Automatic Repeat Request)
- Alternative: Vorwärtsfehlerbehebung (Forward Error Control)
 - k von n Paketen einer Nachricht sind zum Wiederherstellen der Nachricht notwendig
- Kombinationsmöglichkeit von FEC und ARQ
 - Empfänger empfängt $m < k$ Pakete und fordert j Pakete erneut an.
 - Sender wiederholt mindestens j Pakete
 - Übertragungswiederholungen können verschiedene fehlerhafte Pakete reparieren.
- Digitale Fontänen
 - Permanentes Senden von (verschiedenen) redundanten Paketen
- Asynchronous Layered Coding
 - Senden von redundanten Paketen über verschiedene Kanäle
 - Auswahl der Kanäle durch Empfänger anhängig von Stausituation

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

27

Staukontrolle

- TCP-Staukontrolle kann mit einer mathematischen Beziehung approximiert werden.
 - Datenrate wird berechnet als Funktion der
 - Anzahl quittierter Pakete
 - Paketumlaufzeit
 - Retransmission Timeout
 - TCP Friendly Multicast Congestion Control (TFMCC)
- Kombination mit ALC
 - Auswahl von Kanälen durch Empfänger um gleiches Verhalten wie TCP-Staukontrolle zu erreichen.
- Router Paketfilterung
 - Empfänger senden Stauberichte stromaufwärts
 - Router filtern ausgehende Pakete

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

28

IETF WG Reliable Multicast Transport

- Ein einziges zuverlässiges Transport Protokoll für alle möglichen Anwendungsszenarien erscheint nicht sinnvoll.
- daher: Definition von flexibel verwendbaren Komponenten („Building Blocks“)
- spezifizierte Komponenten für
 - Forward Error Control
 - Layered Coding Transport
 - Wave and Equation Based Rate Control
- NACK Gated Reliable Multicast Protocol

SS 02

Torsten Braun (Universität Bern): Netze und Protokolle für das Internet

29