

Verteilte Systeme SS2002:

Gruppenkommunikation

Marcel Waldvogel

Übersicht

- **Paare und Gruppen**
- **Vor-/Nachteile von Gruppen**
- **Protokolle**
 - *Multicast etc.*
- **Applikationen**
 - *News*
 - *Zeit*

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 2

Kommunikationsarten

- **Paarweise**
 - *Client-Server*
 - *Peer-to-Peer*
- **Gruppenweise**
 - *Peer-to-Peer*
 - *Mit Koordinator*

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 3

Client-Server-Paar

- **Anzahl Klienten**
 - *Dedizierte Verbindung*
 - *Polling, Sensor*
 - *Variabler einzelner Klient*
 - *X11 Window Manager*
 - *Konstante Anzahl Klienten*
 - *Messzentrale*
 - *Variable Anzahl Klienten*
 - *Typischer Fall*
- **Koordination beim Server**
 - *Einzelner Worker-Prozess*
 - *Mehrere: Gegenseitiger Ausschluss*

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 4

Peer-to-Peer-Paar

■ Beispiel

- Primary-/Backup-Server
- Protokolle: TCP, ...

■ Gegenseitiger Ausschluss nötig?

- Koordination?
- Zuverlässigkeit?

■ Authentisierung

- IP-Adresse
- Kryptografisch

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 5

Wieso Gruppen?

■ Vorteile

- **Metcalf's Law:**
Der Nutzen von Kommunikationsdiensten steigt quadratisch mit der Anzahl Benutzer (und damit proportional zu den möglichen Verbindungen)
 - Geschlossene vs. offene Standards; Beispiel WWW
- The Computational Grid
- Skalierung!

■ Nachteile

- Koordination
- Netzwerk (Bandwidth, Delay, Loss)
- Mehrere Rechner (Administration, Ausfall, ...)

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 6

Kategorien von Gruppen

■ Welche Probleme tauchen auf bei mehreren

- Klienten?
- Servern?
- Peers?

■ Abhängig von Applikation?

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 7

Skalierung

■ Parallelität

- Analog zu Parallelrechnern

■ Enge vs. lose Koppelung

- distributed.net
 - Client-Server
- Datenbanken
 - Peer

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 8

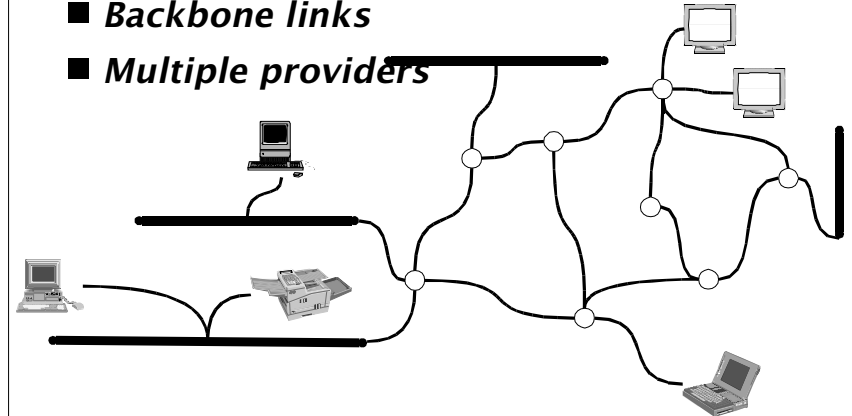
Mechanismen

- ***cast**
 - Unicast
 - Broadcast
 - Multicast
 - Anycast
- **Ringe**
- **Bäume**
- **Netze**

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 9

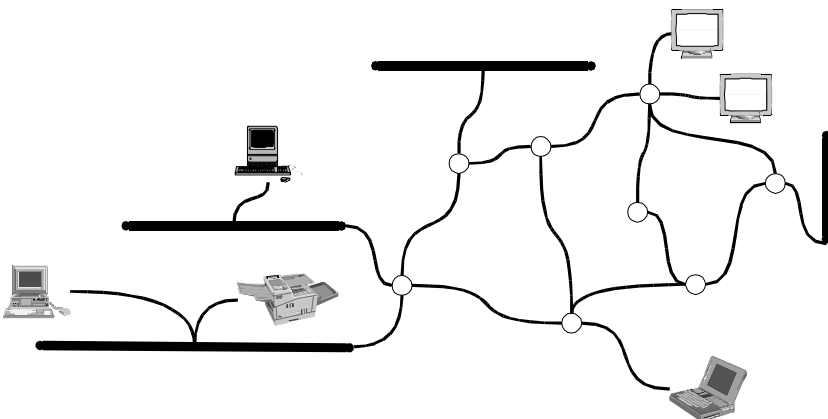
Internet: A Peek Inside

- **Computers and LANs**
- **Backbone links**
- **Multiple providers**



Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 10

Single Server and Single Client



Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 11

Many Clients

- **Issues**
 - > 50 Million Hosts
 - »Slashdot Effect«
 - Popular Broadcasting Events
- **Effects**
 - Server overload
 - Network overload close to the server

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 12

General Solutions

- **Raw power**
 - *Bigger Servers*
 - *Faster Network*
- **Decentralization**
 - *More Servers*
- **Network support**
 - *Caches*
 - *Broadcast/Multicast*
- **New paradigms**

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 13

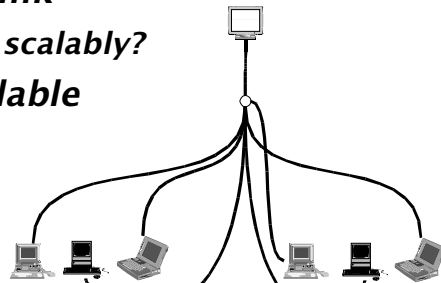
Raw Power

- **Bigger, faster computers and networks**
- **Split the problem**
 - *Distribute the requests*
 - *Client side: Randomly pick a server from a list*
 - *Server side: Virtual server, distribute requests to real servers*
- **Costly**
- **Does not scale well**
- **Servers and network bandwidth need to grow linearly**

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 14

Multicast Problems

- **Packet loss and retransmission**
 - *»Sender implosion«*
 - *Guaranteeing delivery*
- **Fair rate at each link**
 - *How to determine scalably?*
- **Make routing scalable**



Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 15

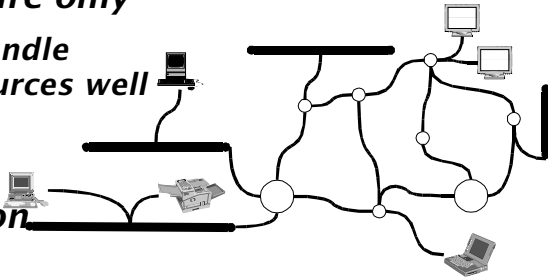
Distribute Servers

- **Have servers all over the world**
- **Network bandwidth no longer a bottleneck**
- **Better latency**
- **Issues**
 - *Machines still need to grow linearly*
 - *Synchronization*
 - *Management nightmare*
 - *How do clients find the closest server?*

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 16

Caching

- **Well-known and widely used for WWW**
- **Static content only**
- **User tracking hard (e.g., shopping basket)**
- **Tree structure only**
 - *Does not handle multiple sources well*
- **Needs manual configuration**



Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 17

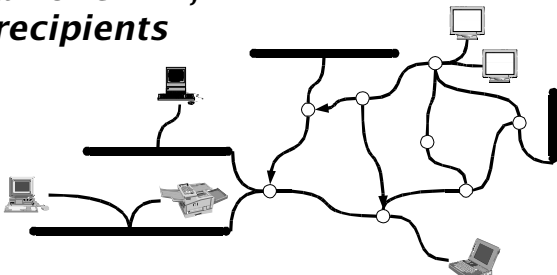
Broadcast

- **Well-known from TV and Radio**
- **Frequencies (=bandwidth) used in entire reception area**
 - *Independent of interested receivers*
- **Does not scale to global reception of many stations**
- **Broadcast: »dumb« air waves**
- **Network: »intelligent« routers**
 - *Improvements worth the cost?*

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 18

Introducing: Multicast

- **Unicast: Every router sends data out on a single link to get it closer to the single destination**
- **Multicast: Data goes out on more than one link, if multiple recipients exist**



Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 19

Multicast

- **Perfect solution?**
 - *Defined for Internet since 1991*
 - *Extremely limited availability*
- **Routing protocol expensive**
 - *Routing traffic*
 - *Router memory*
- **ISPs are afraid**
 - *Data traffic*
 - *Reliability*
 - *Charging*

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 20

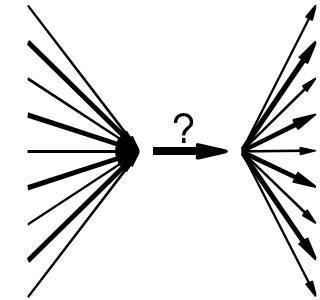
Internet Policy

- **Fair bandwidth sharing**
 - No enforcement
- **Routers still relatively dumb**
 - Cost/performance
 - Only tries to forward packets
 - No retransmits
 - No information processing
 - Overload notified as packet loss

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 21

Congestion Control

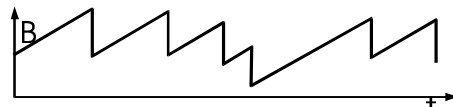
- **Why fairness?**
- **How to achieve?**
- **How to find out about fair share?**



Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 22

Congestion Control: TCP/IP

- **Router »dumb«**
- **Router provides random packet losses on overutilized links**
 - Receiver reports loss to sender
 - Sender reduces transmission rate at every loss
- **Large flows see more losses**
- **To use available bandwidth, senders increase data rate slowly**



Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 23

Congestion Control: Multicast Issues

- **Worst case congestion as basis**
 - As with unicast
 - Policy cut-off
- **Packet loss feedback not scalable**
 - Implosion
- **Drop-to-zero problem**
 - Loss rate, not loss events

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 24

SRM/CC

- **Only small number (~1) of receivers provide feedback**
- **Dynamic election process**
 - *Worst candidate*
 - *Piggybacked on retransmission request*
 - *Aggregated*
 - *Probabilistic*
 - *Most losses*
 - *Low-pass filter*

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 25

Anycast

- **IPv6**
- **Global Internet Anycast**
- **Adresse eines möglicherweise replizierten Dienstes**
- **Routinginformation zur Lokalisierung**

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 26

Applikationen

- **Zeitsynchronisation (NTP)**
- **NNTP**
- **Datensynchronisation**
- **AFS**

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 27

Zeit

- **Synchrone Zeit wichtig**
- **Lichtgeschwindigkeit endlich**
 - *1GHz ~ 1ps ~ 20cm*
 - *1kHz ~ 1ms ~ 200km*
- **Global Position System (GPS)**
- **Network Time Protocol (NTP)**
 - *Netzwerke weder deterministisch noch symmetrisch*
 - *Frequenz und Phase*
- **Ordnung**

Verteilte Systeme SS 2002, Marcel Waldvogel, IBM ZRL+ETHZ, 28.05.2002, 28

NNTP

- ***Network News Transport Protocol, 1986***
 - *Globales Diskussionsforum*
 - *Replikation*
 - *Grosse Datenmengen*
- ***Redundantes Netz***
- ***IHAVE/SENDME mit Message-IDs***
 - *Ineffizient*

Massensynchronisation

- ***Effizienz steigern, aber wie?***
- ***Annahmen:***
 - *Viele Nachrichten*
 - *Relativ wenige Quellen (Hunderte)*

Andrew File System (AFS)

- ***Verteiltes Dateisystem (CMU, IBM, Open)***
- ***Baum von (replizierten) Server***
 - *2PC*
- ***Aggressives Caching der Clienten***
 - *Callbacks mit Limiten*
 - *Nach Dateimodifikation: Client als Server*

Weitere Applikationen