CoNe Freiburg

IIF INSTITUT FÜR INFORMATIK FREIBURG

Algorithmen und Datenstrukturen für Peer-to-Peer-Netzwerke

Christian Schindelhauer
Technische Fakultät
Rechnernetze und Telematik
Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

CoNe Freiburg

Algorithmen und Methoden für Peer-to-Peer-Netzwerke

- Entwicklung
  - Napster
  - Gnutella
  - DHT
  - CAN
- Graphstrukturen
  - Chord
  - Pastry, Viceroy, Distance-Halving
- Durchsatz
  - IP Multicast, Splitstream
  - Bittorrent, Spieltheorie
- Netzwerk-Kodierung
  - Network-Coding, Pair-Coding, Tree Network Coding
- Zusammenfassung und Ausblick

UNI FREIBURG

CoNe Freiburg

1996

- Anzahl der Promotionen in Informatik an der Universität Lübeck

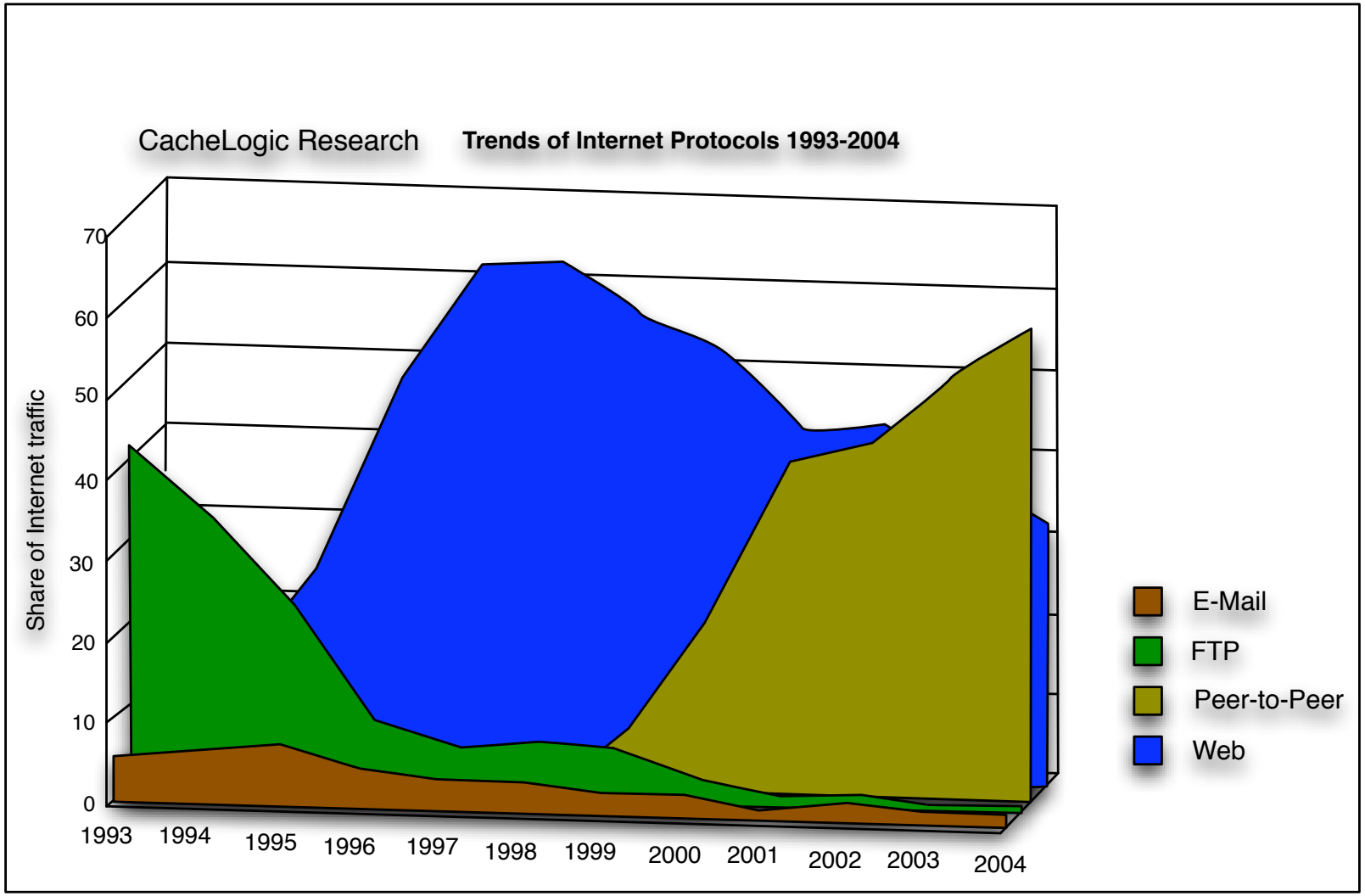
1

- Anzahl der Peer-to-Peer-Netzwerke weltweit:

0

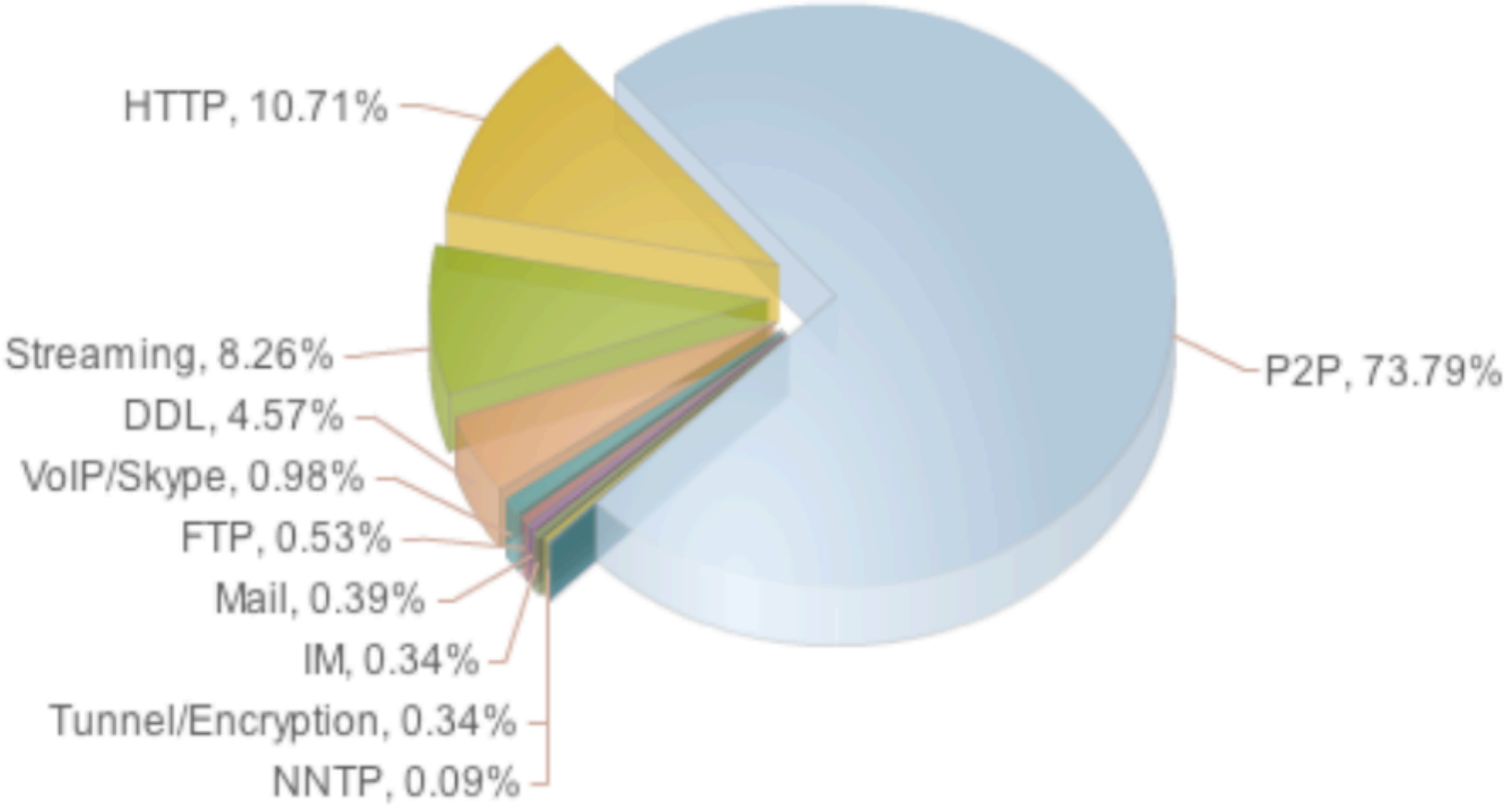
UNI FREIBURG

Global Internet Traffic Shares 1993-2004



CoNe Freiburg

P2P Share Germany 2007

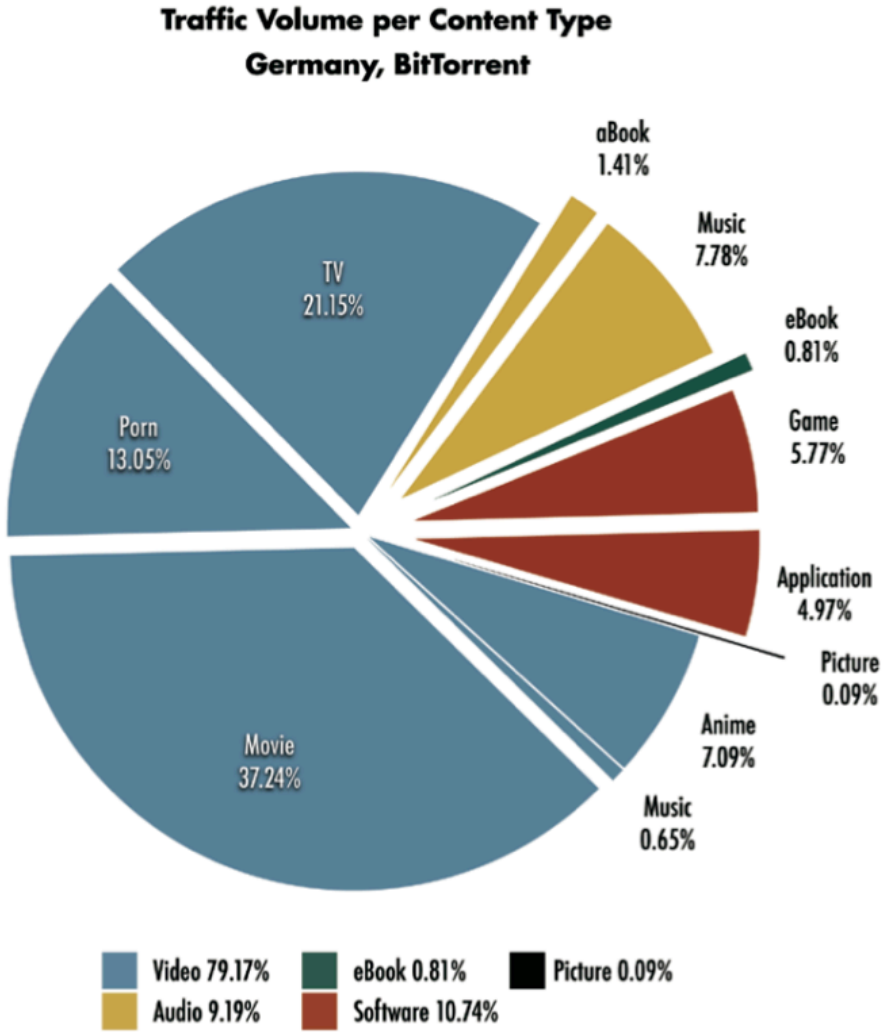


Quelle: Ipoque 2007

UNI FREIBURG

CoNe Freiburg

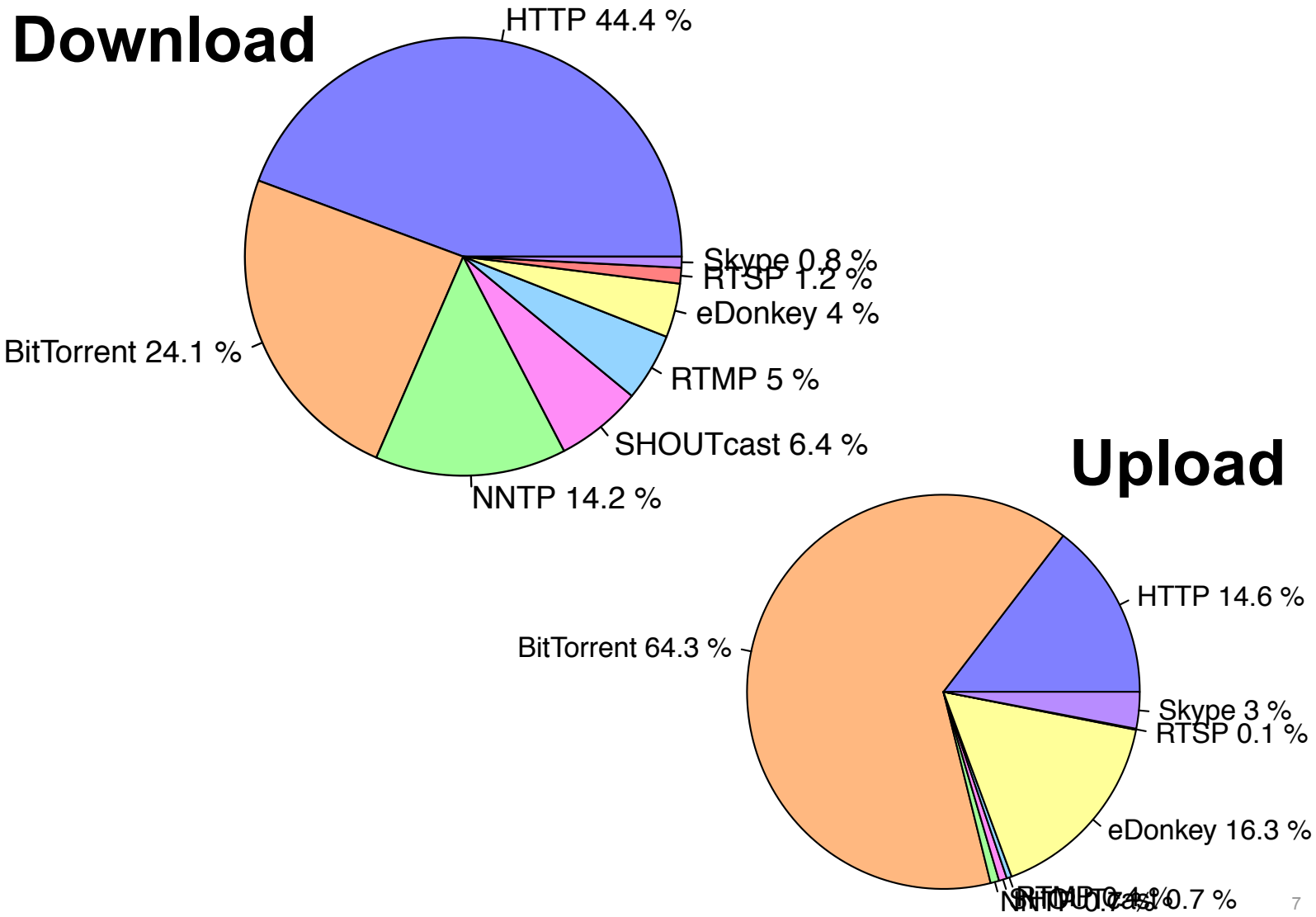
What Germans Download 2007 by Volume



Quelle: Ipoque 2007

CoNe Freiburg

Internet Traffic eines Deutschen ISP August 2009



CoNe Freiburg

Meilensteine P2P Systeme

- Napster 1999-2000
  - Filesharing, nur rudimentäres P2P
- Gnutella 2000
  - 1. echtes P2P-Netzwerk
- Edonkey 2000
  - Mehr Filesharing als P2P
- FreeNet 2000
  - Anonymisiertes P2P-Netzwerk
- FastTrack 2001
  - KaZaa, Morpheus, Grokster
- Bittorrent 2001
- Skype 2003
  - VoIP (voice over IP), Chat, Video

UNI FREIBURG

CoNe Freiburg

Milestones Theorie

- Distributed Hash-Tables (DHT) (1997)
  - Ziel: Lastbalancierung für Web-Server
- CAN (2001)
  - DHT-Netzwerk-Struktur
- Chord (2001)
  - Erstes effiziente P2P-Netzwerk
  - Logarithmische Suchzeit
- Pastry/Tapestry (2001)
  - Effizientes verteiltes P2P-Netzwerk unter Verwendung des Plaxton-Routing
- Und viele andere Ansätze
  - Viceroy, Distance-Halving, Koorde, Skip-Net, P-Grid, ...
- In den letzten fünf Jahren:
  - Network Coding for P2P
  - Game theory in P2P
  - Anonymity, Security

UNI FREIBURG

CoNe Freiburg

Was ist ein P2P-Netzwerk

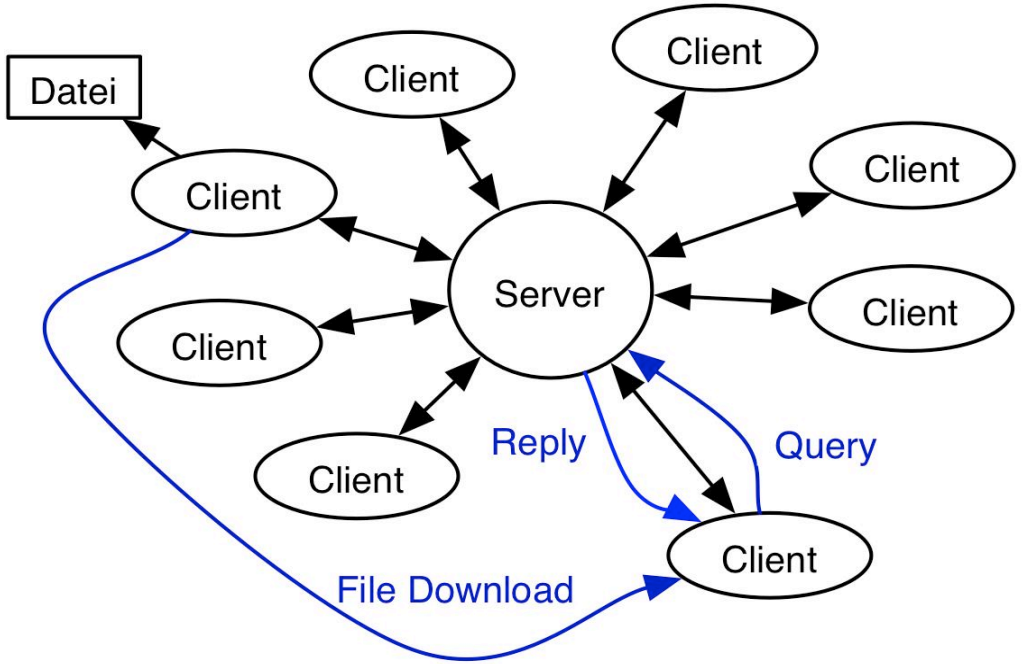
- Was ist P2P **NICHT**?
  - Ein Client-Server network
- Etymologie: peer
  - lateinisch: par = gleich
  - Standesgleich
  - P2P, Peer-to-Peer: Beziehung zwischen gleichwertigen Partnern
- Definition
  - Ein Peer-to-Peer Network ist ein Kommunikationsnetzwerk im Internet
    - ohne zentrale Kontrollel
    - mit gleichwertigen, unzuverlässigen Partnern

UNI FREIBURG

CoNe Freiburg

Napster

- Shawn (Napster) Fanning
  - Juni 1999 Beta-Version von Napster
- Ziel: File-sharing-System
  - Tatsächlich: Musik-Tauschbörse
  - Herbst 1999 Download des Jahres
- Client-Server-Struktur

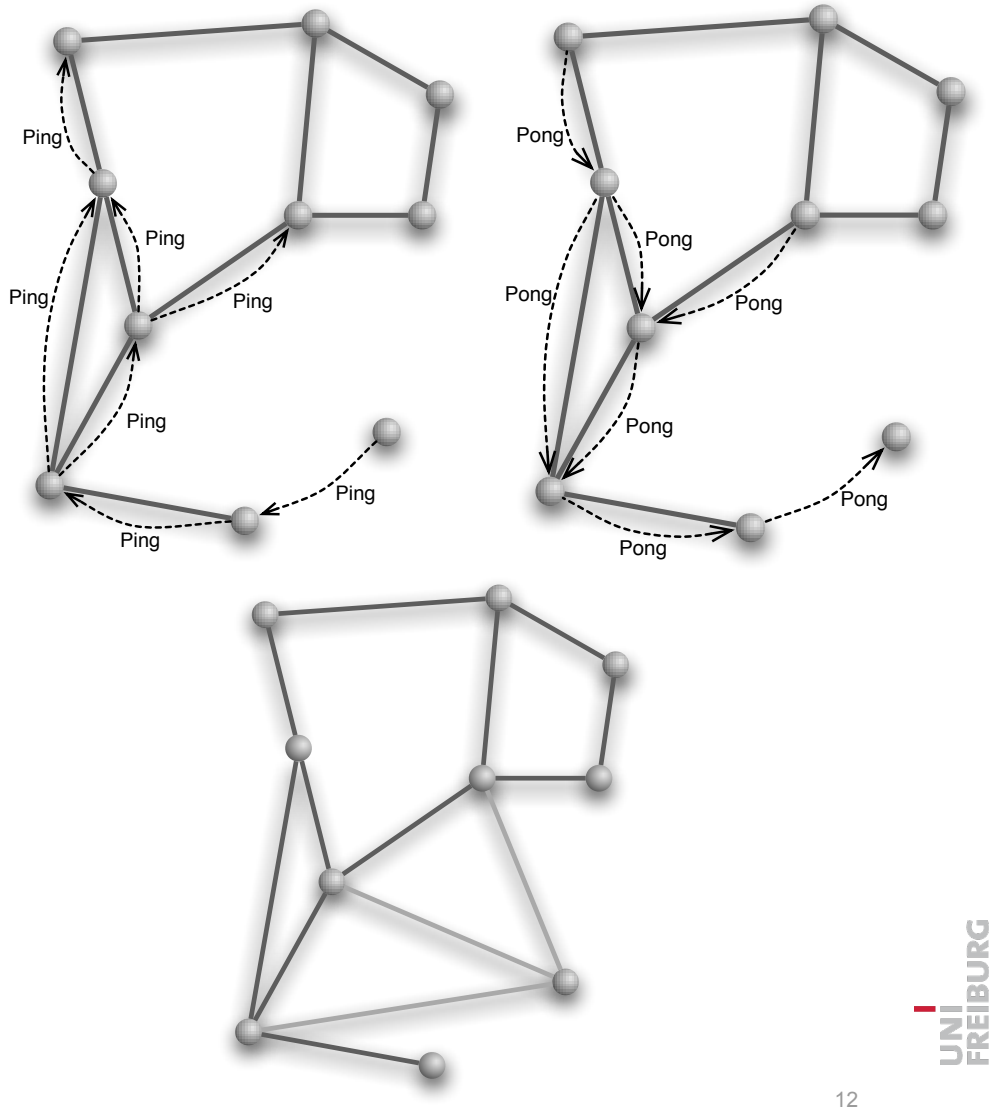


UNI FREIBURG

CoNe Freiburg

Gnutella

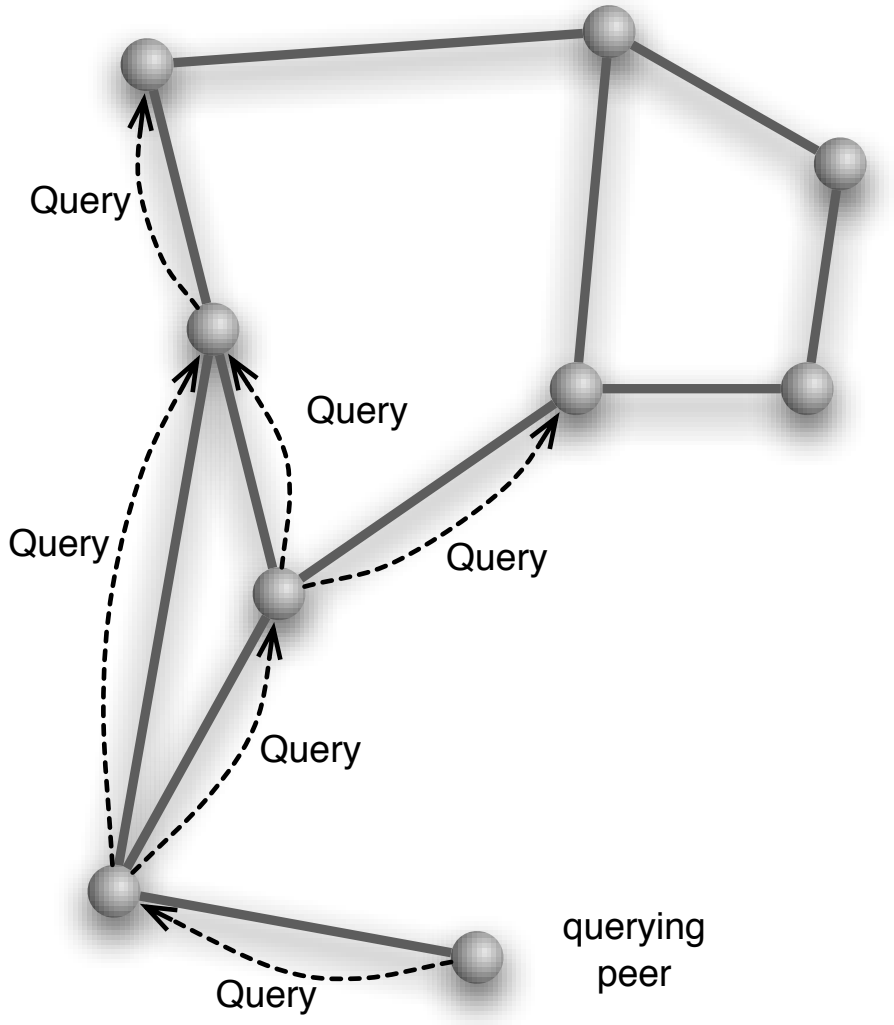
- Gnutella
  - März 2000 von Justin Frankel und Tom Pepper
- File-Sharing-System
  - Ziel wie Napster
  - aber völlig ohne zentrale Strukturen
- Struktur und Suche
  - Aufbau durch Ping/ Pong-Nachrichten an Nachbarn
  - Suche durch tiefenbeschränktes Fluten



UNI FREIBURG

CoNe Freiburg

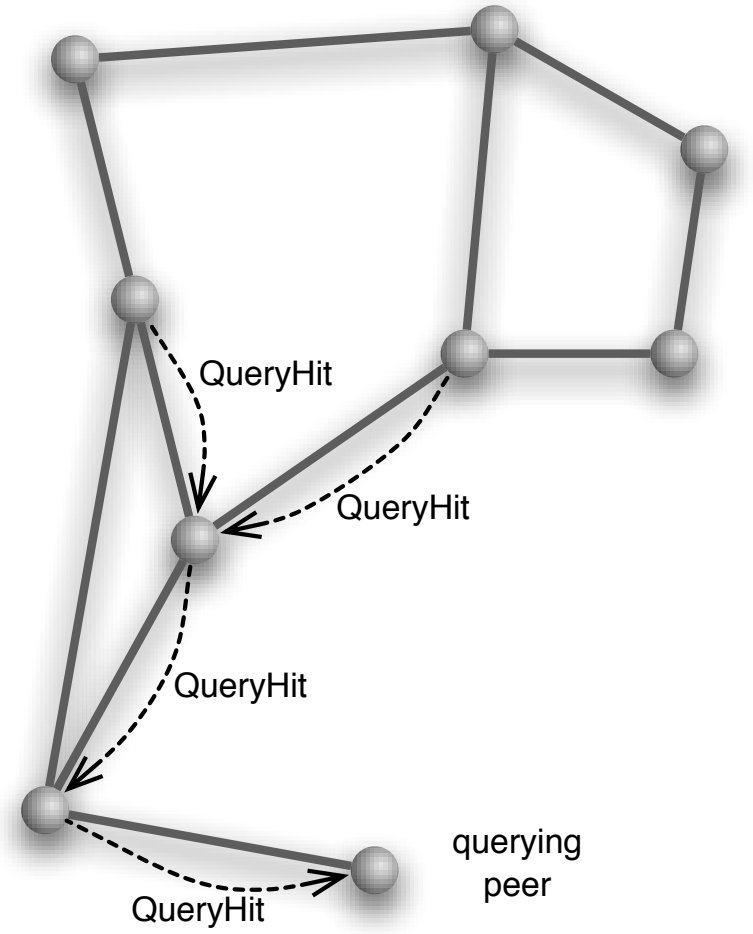
Gnutella — Suche



UNI FREIBURG

CoNe Freiburg

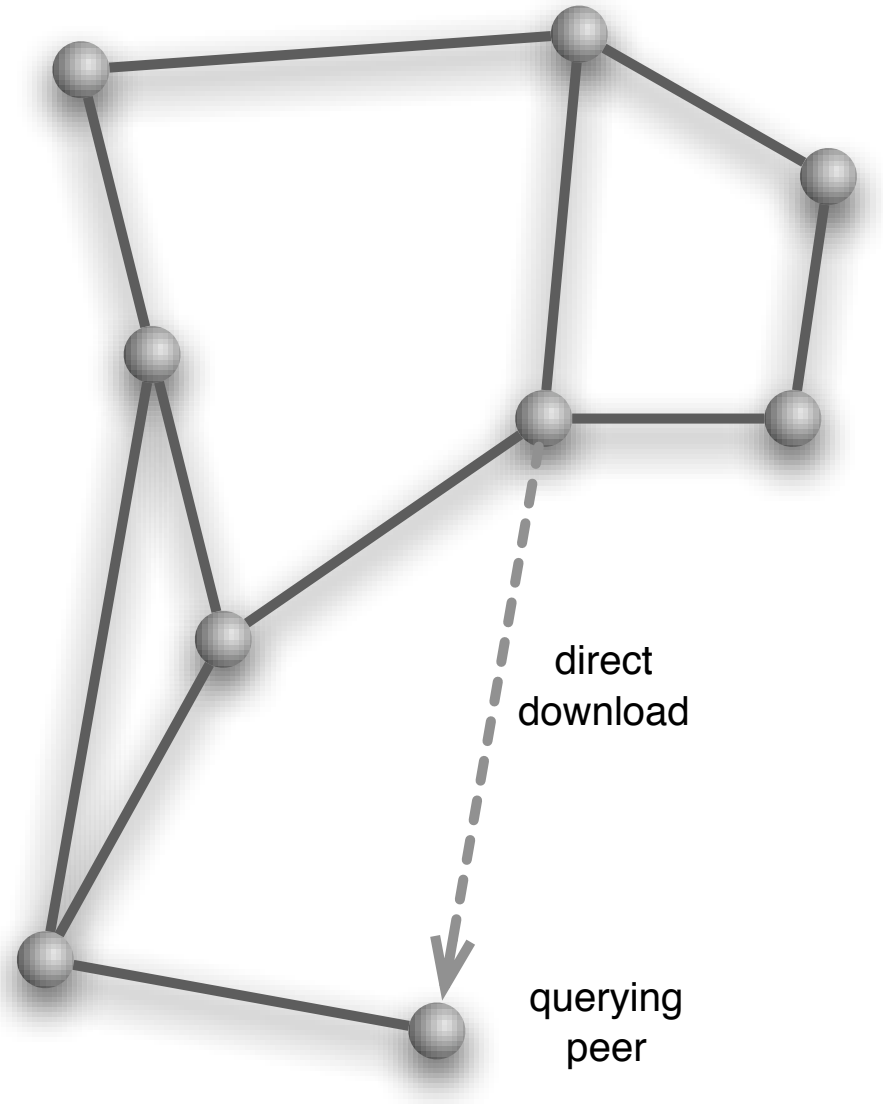
Gnutella — Suche



UNI FREIBURG

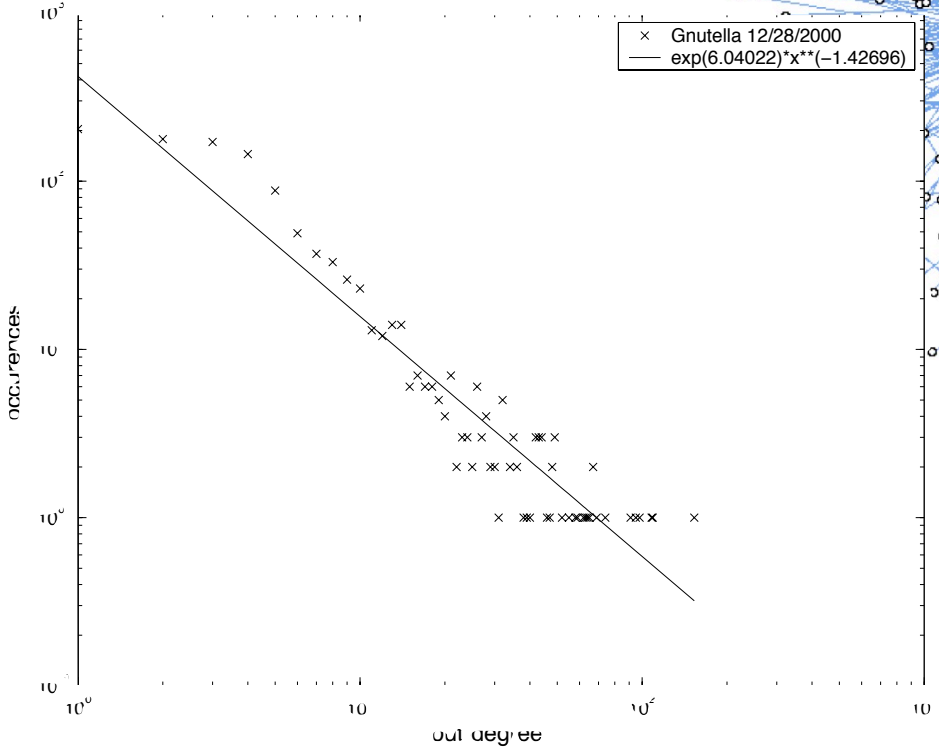
CoNe Freiburg

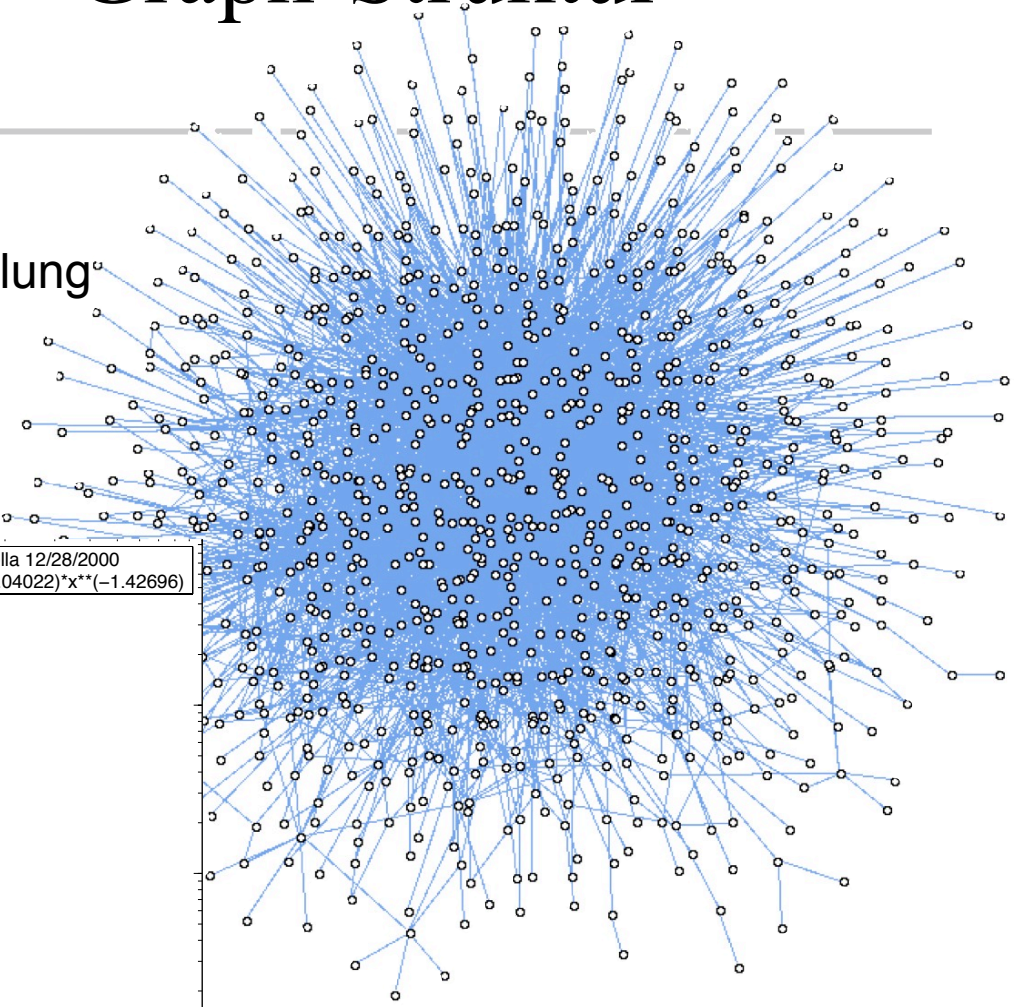
Gnutella — Suche



Gnutella — Graph-Struktur

- Zufällig
- Grad unterliegt Pareto-Verteilung
- Unkontrolliert und robust
- Ineffizient und ineffektiv

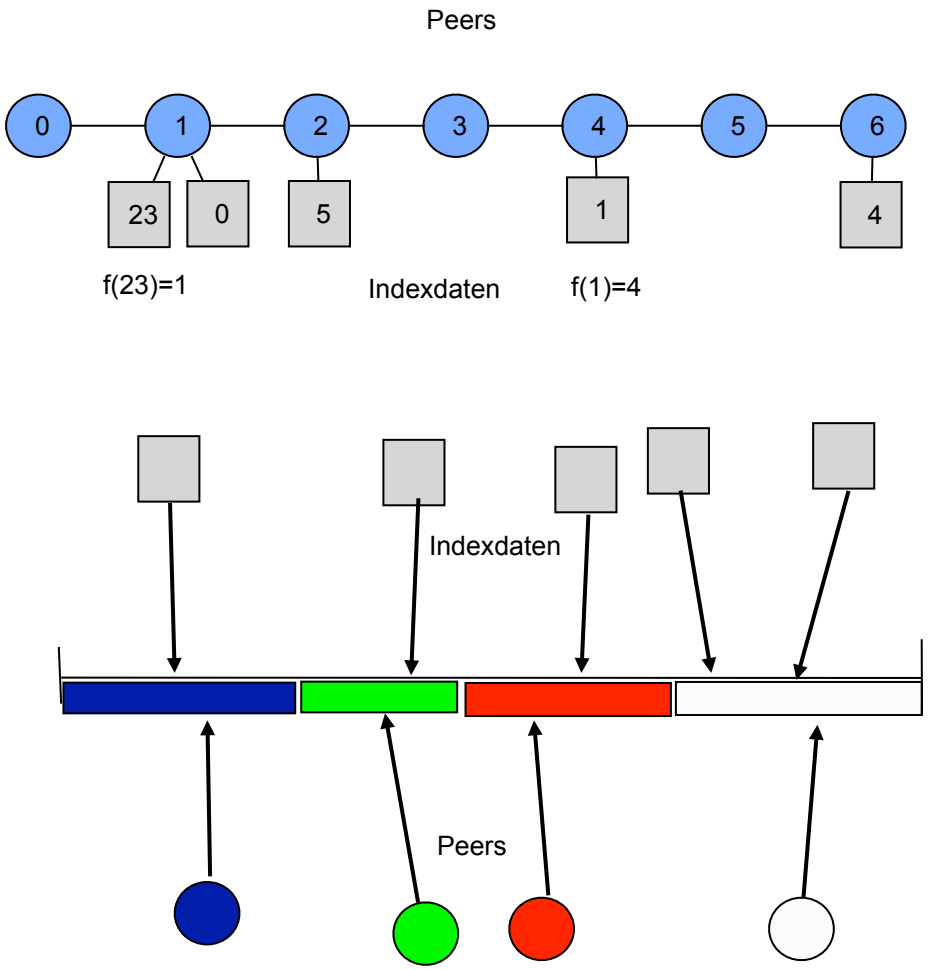




Gnutella Schnappschuss 2000

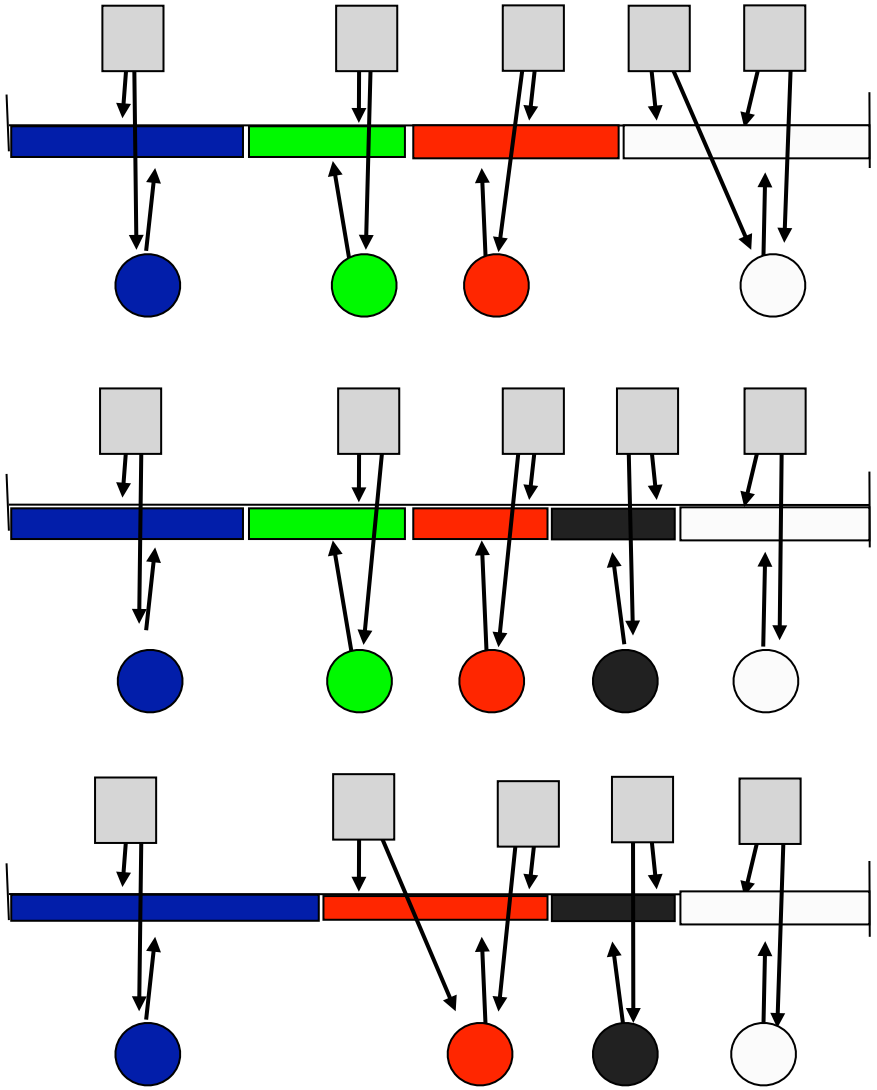
Distributed Hash-Table (DHT)

- Hash-Tabellen
  - nicht praktikabel in P2P
- Verteilte Hash-Tabellen
  - *Consistent Hashing and Random Trees: Distributed Caching Protocols for Relieving Hot Spots on the World Wide Web*, Karger, Lehman, Leighton, Levine, Lewin, Panigrahy, STOC 1997
- Daten
  - werden ge*hash*t und nach Bereich den Peers zugeordnet
- Peers
  - werden an eine Stelle ge*hash*t und erhalten Bereiche des Wertebereichs der Hashfunktion zugeteilt



CoNe Freiburg

Einfügen und Löschen in einer DHT

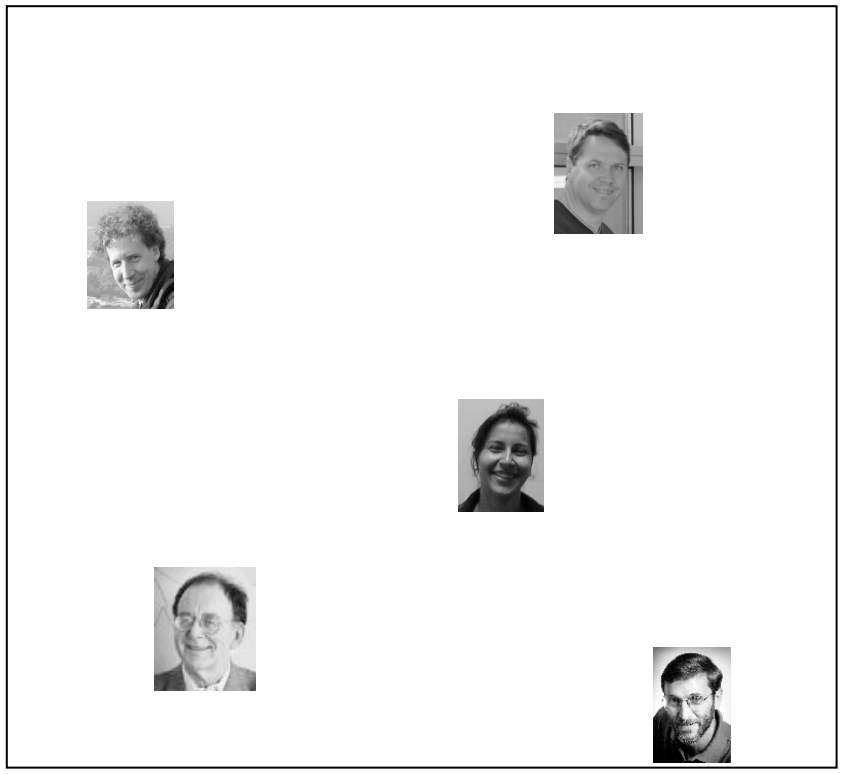


UNI FREIBURG

CoNe Freiburg

Content Addressable Network (CAN)

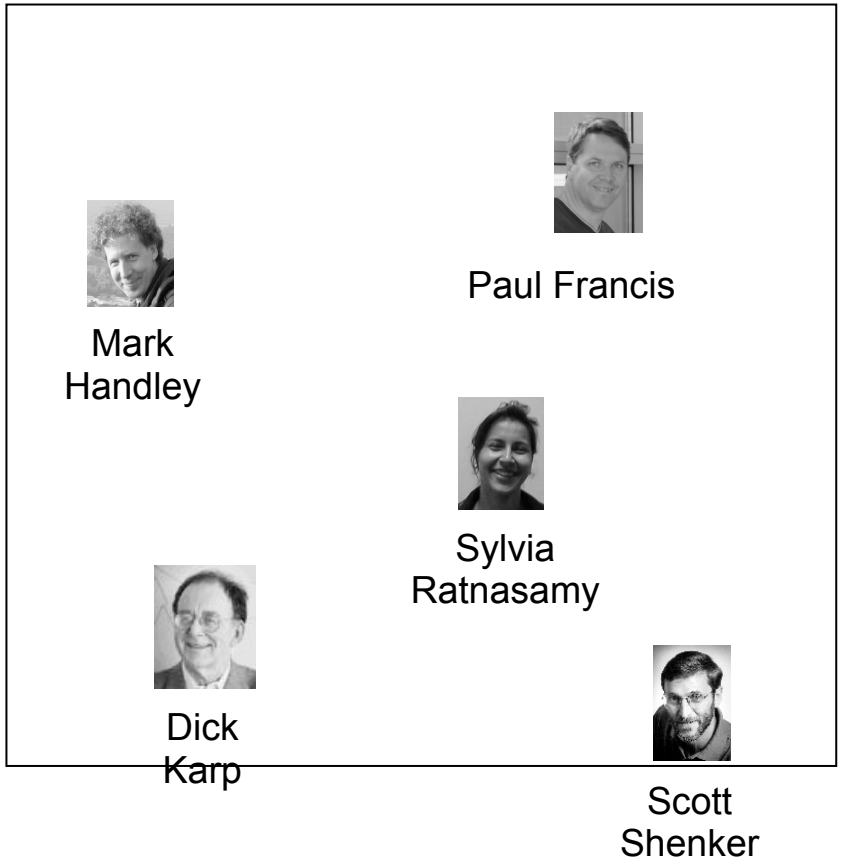
- Dateien werden in durch (zweiwertige)-Hash-Funktion in das Quadrat abgebildet



UNI FREIBURG

CoNe Freiburg

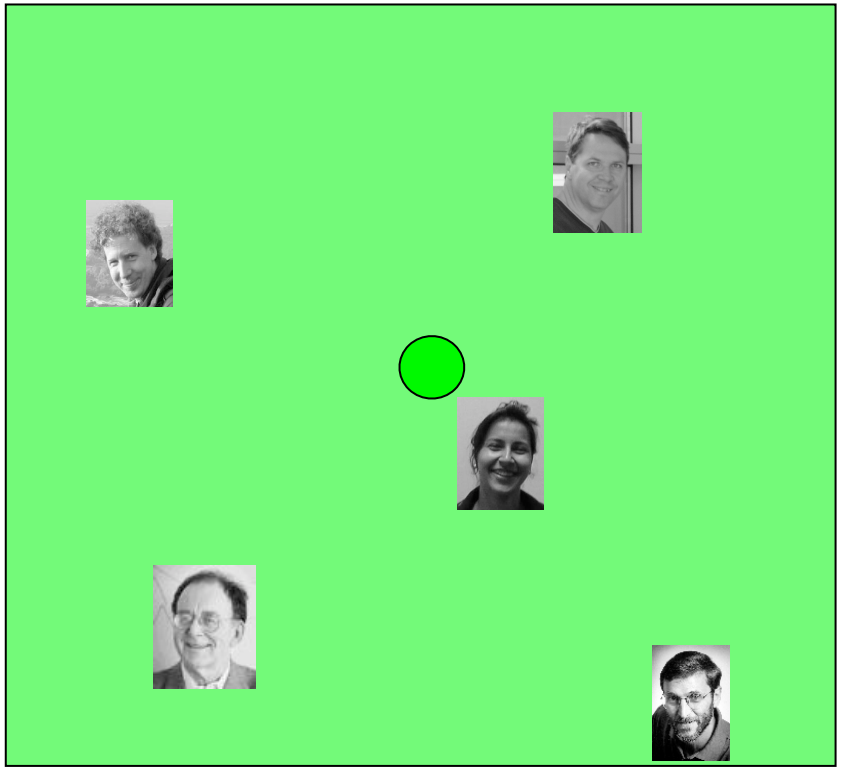
Content Addressable Network (CAN)



UNI FREIBURG

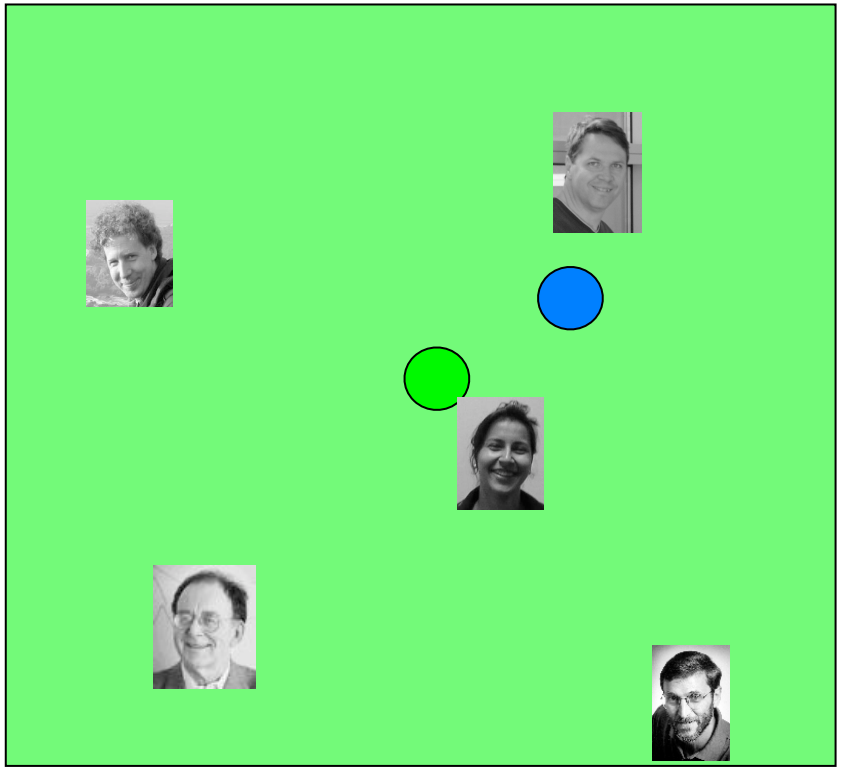
CoNe Freiburg

Content Addressable Network (CAN)



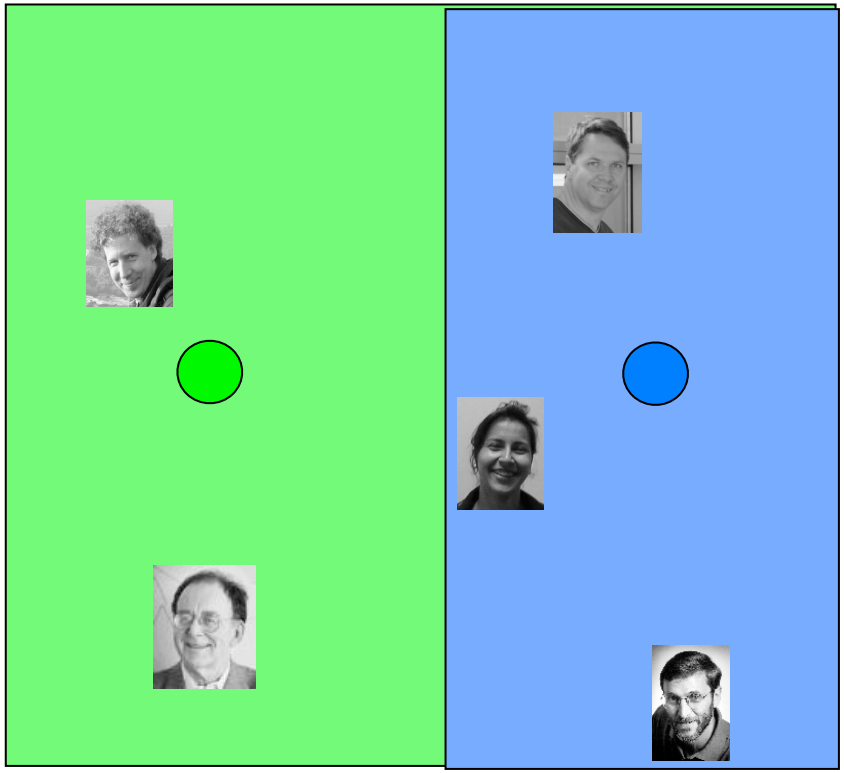
CoNe Freiburg

Content Addressable Network (CAN)



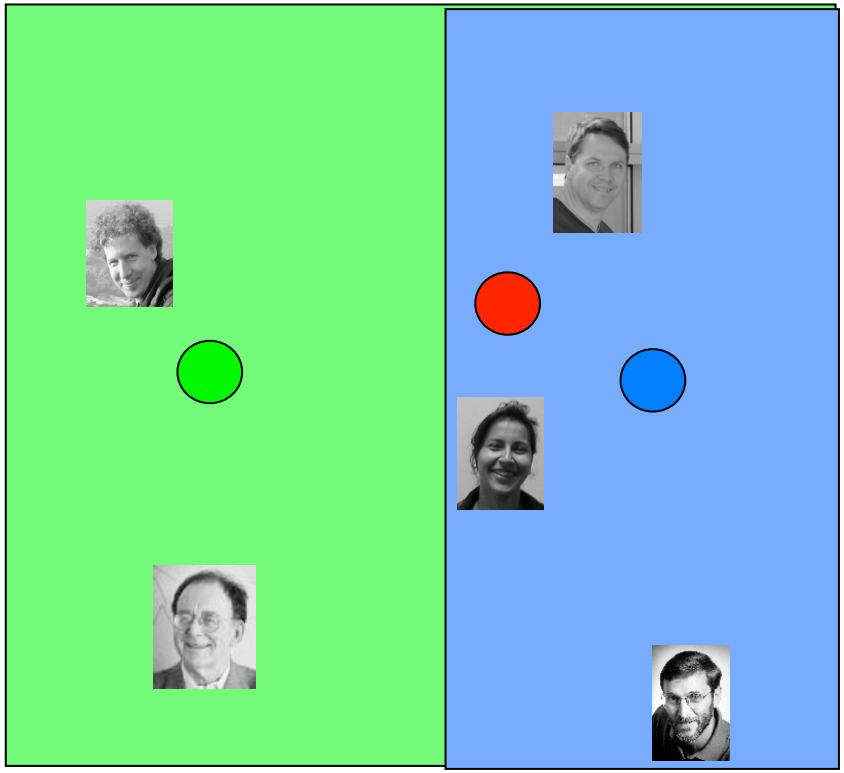
CoNe Freiburg

Content Addressable Network (CAN)



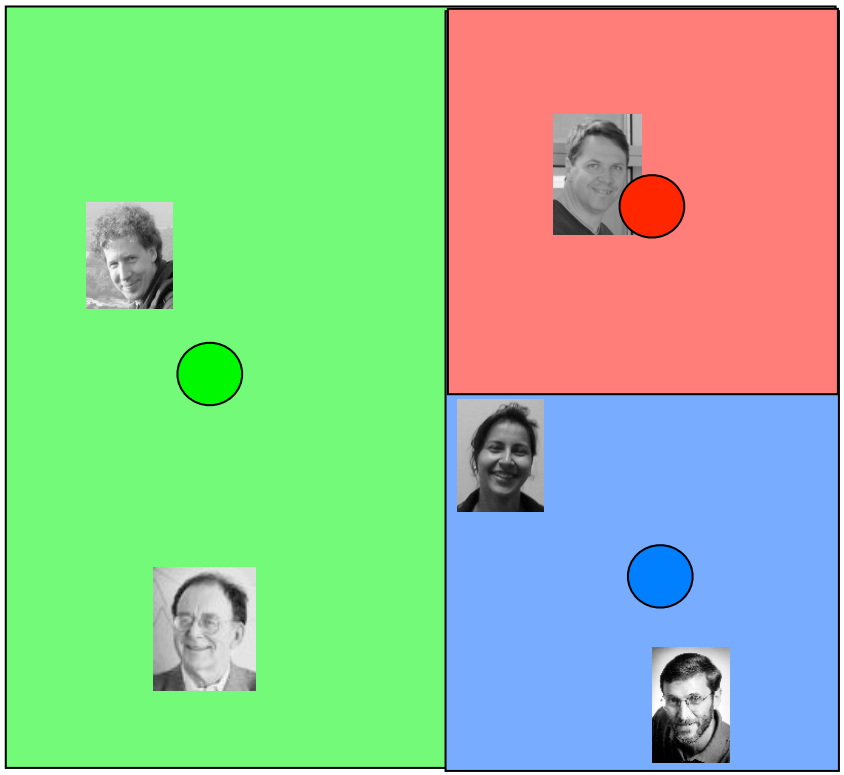
CoNe
Freiburg

Content Addressable Network (CAN)



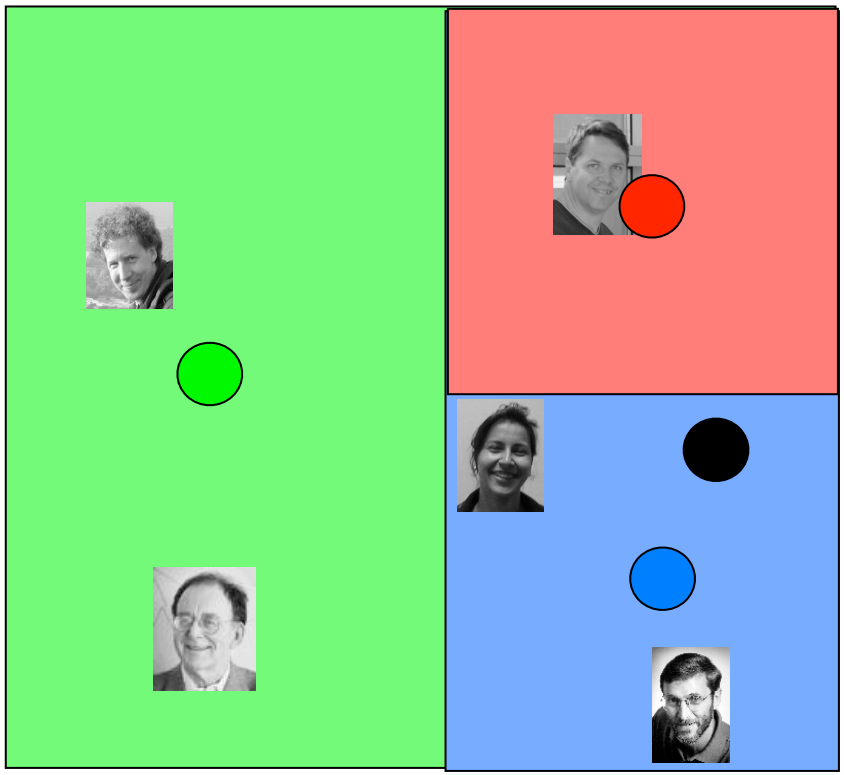
CoNe Freiburg

Content Addressable Network (CAN)



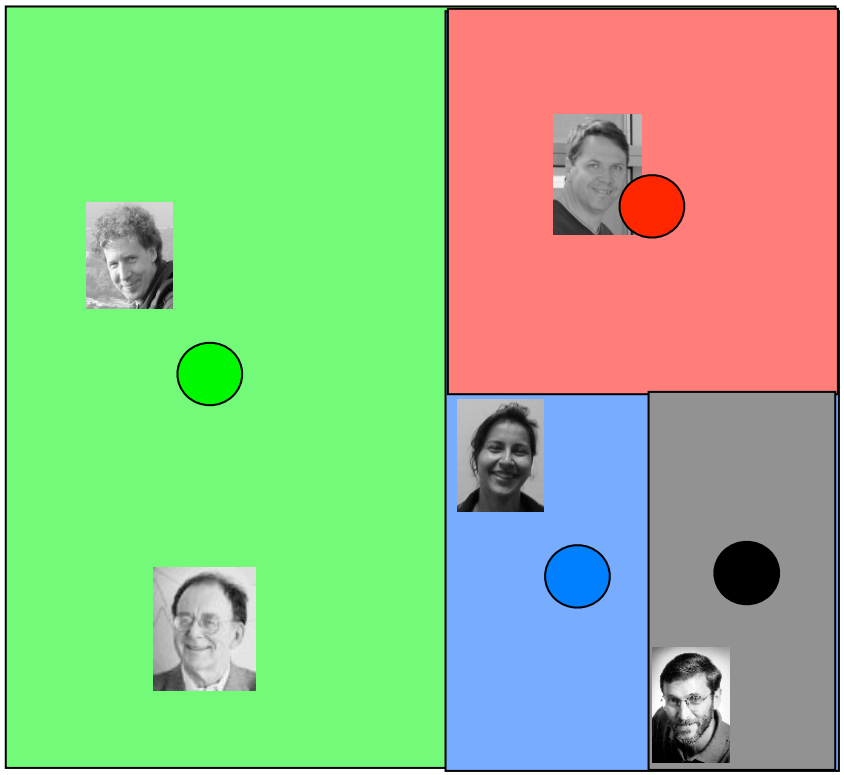
CoNe Freiburg

Content Addressable Network (CAN)



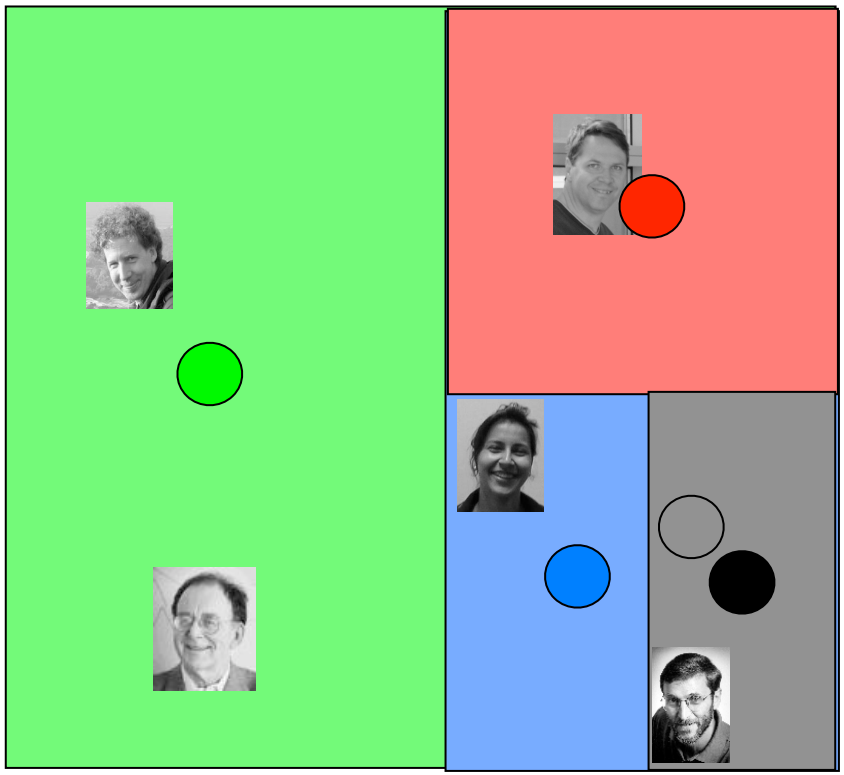
CoNe Freiburg

Content Addressable Network (CAN)



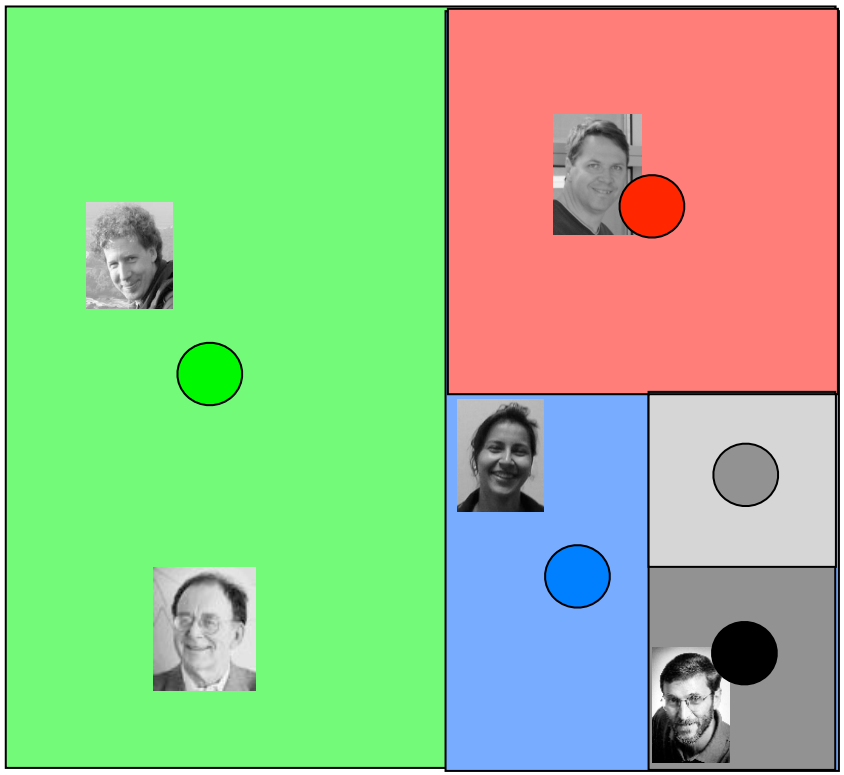
CoNe Freiburg

Content Addressable Network (CAN)



CoNe Freiburg

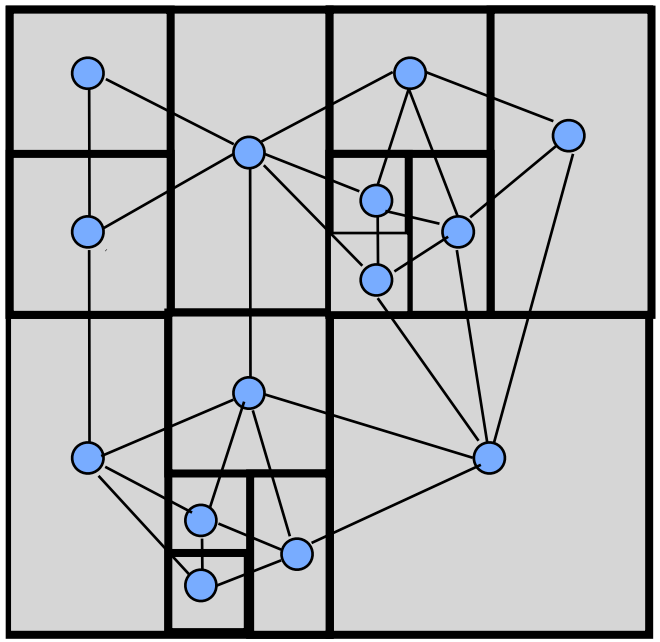
Content Addressable Network (CAN)



CoNe Freiburg

Netzwerk-Struktur von CAN

- Jeder Peer hat Verbindungen zu seinen geometrischen Nachbarn
- Erwarteter Netzwerkgrad ist proportional zur Dimension d
- Netzwerkdurchmesser: $O(n^{1/d})$
- Erwarteter Anteil der Datenmenge ist O(1/n)

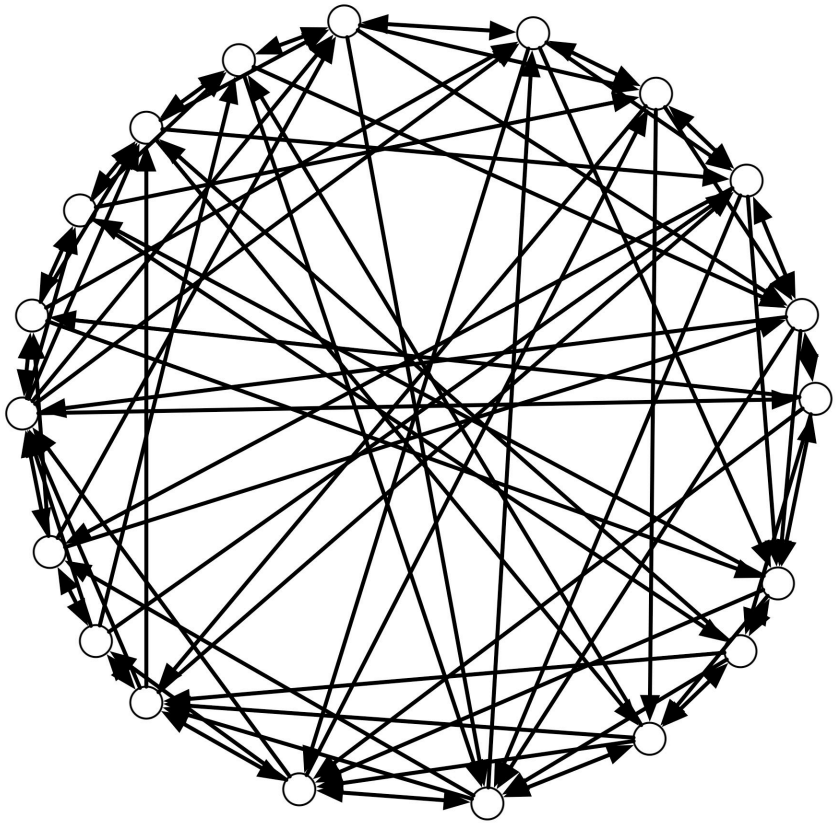


UNI FREIBURG

CoNe Freiburg

Chord

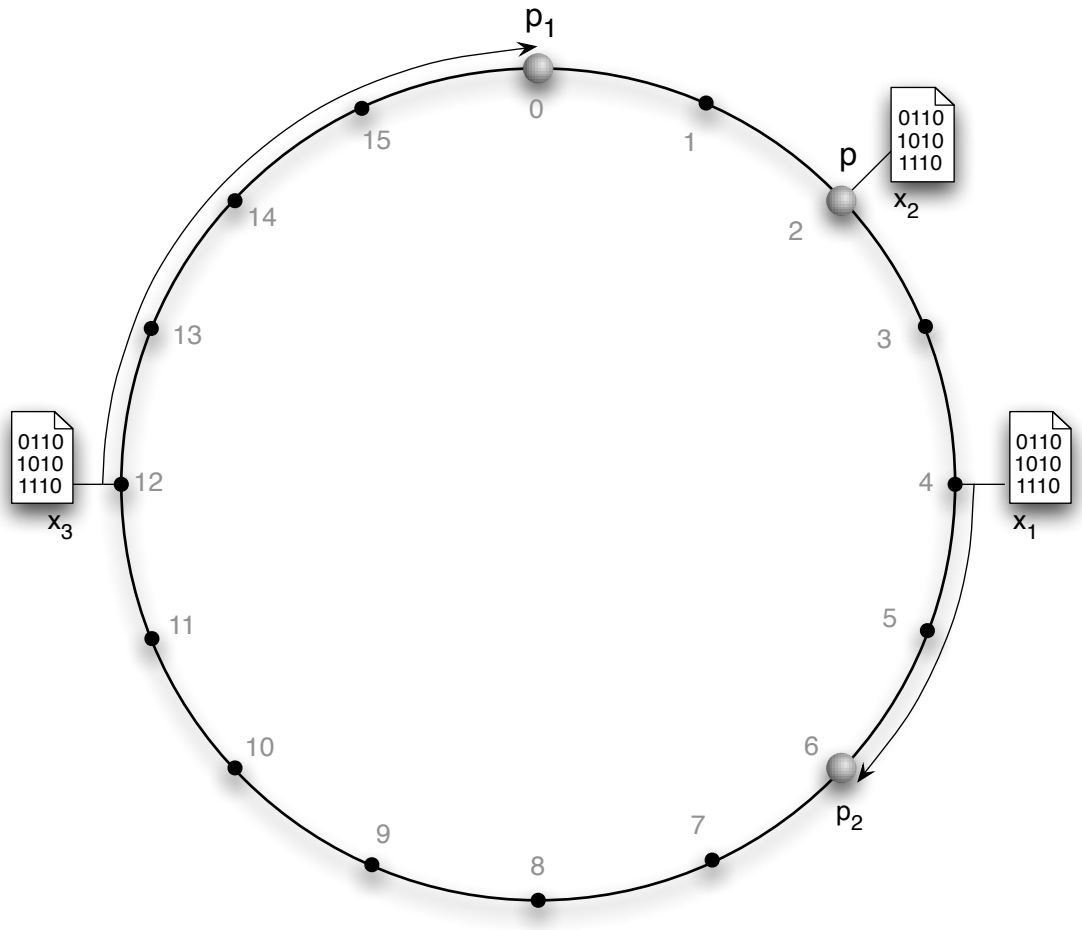
- Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek and Hari Balakrishnan (2001)
- DHT mit logarithmischer Suche



UNI FREIBURG

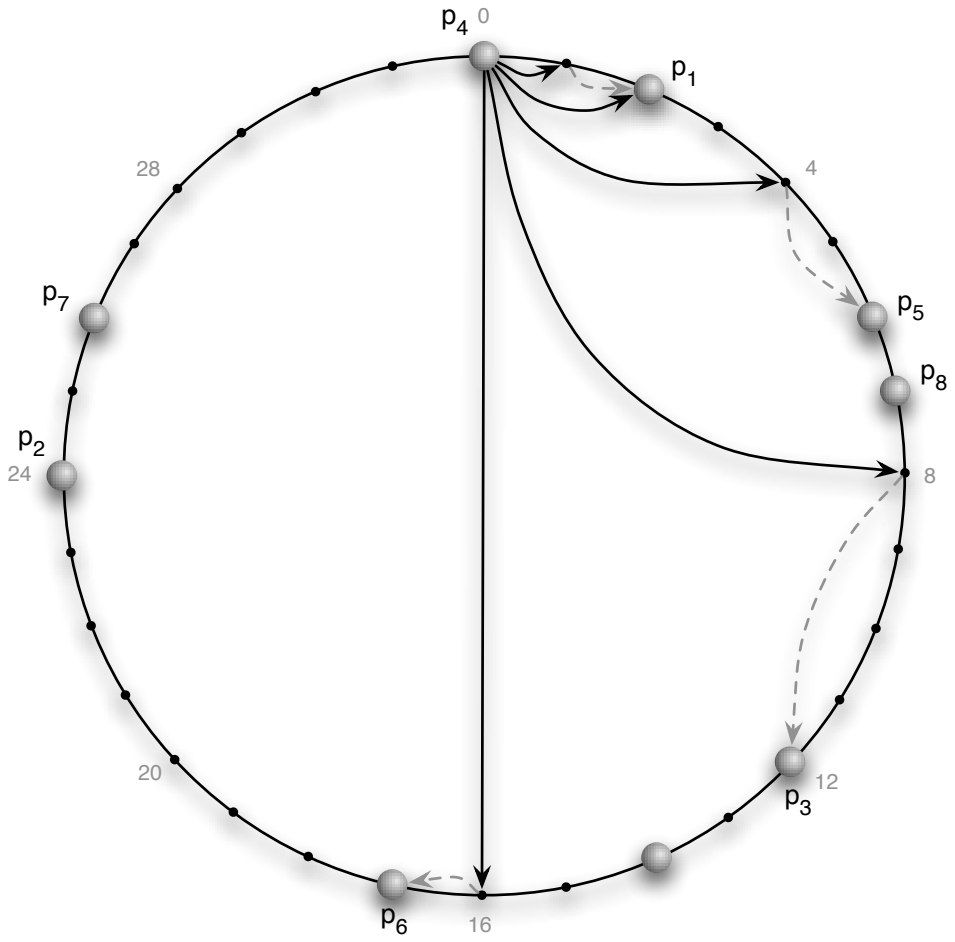
CoNe Freiburg

DHT in Chord



CoNe Freiburg

Zeiger-Struktur in Chord

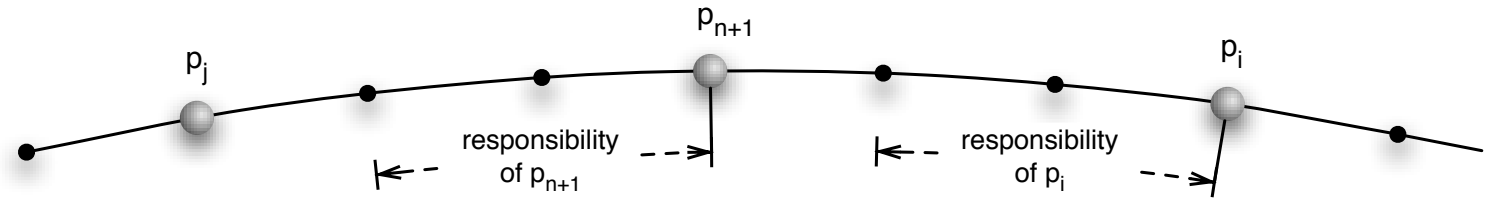


UNI FREIBURG

CoNe Freiburg

Balance in Chord

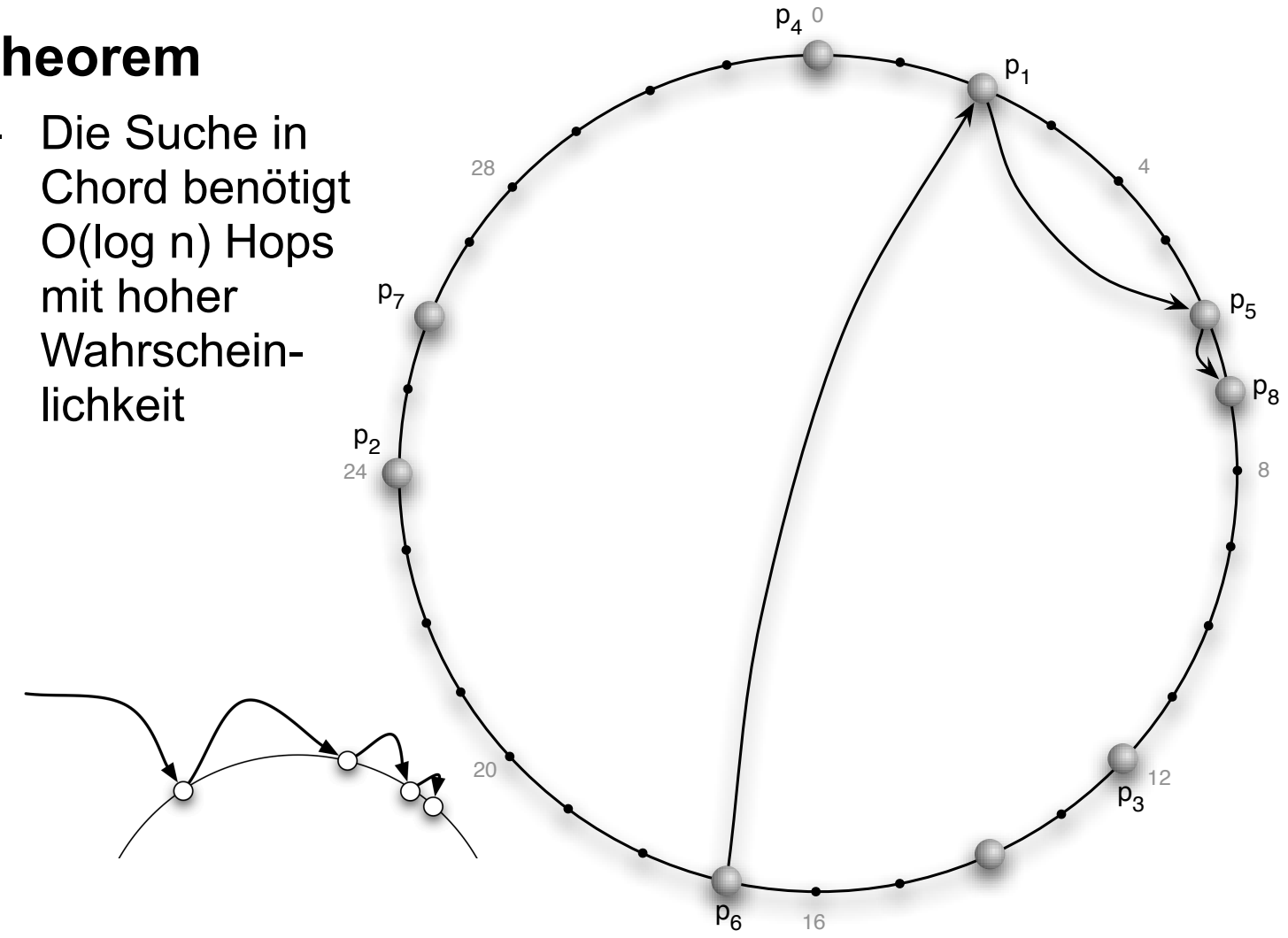
- **Theorem**
  - Jeder der n Peers in Chord speichert von den k Dateneinträgen höchstens O(k/n log n) Einträge mit hoher Wahrscheinlichkeit
  - Wenn ein Peer das Netzwerk betritt oder verlässt, wird höchstens diese Menge bewegt.



UNI FREIBURG

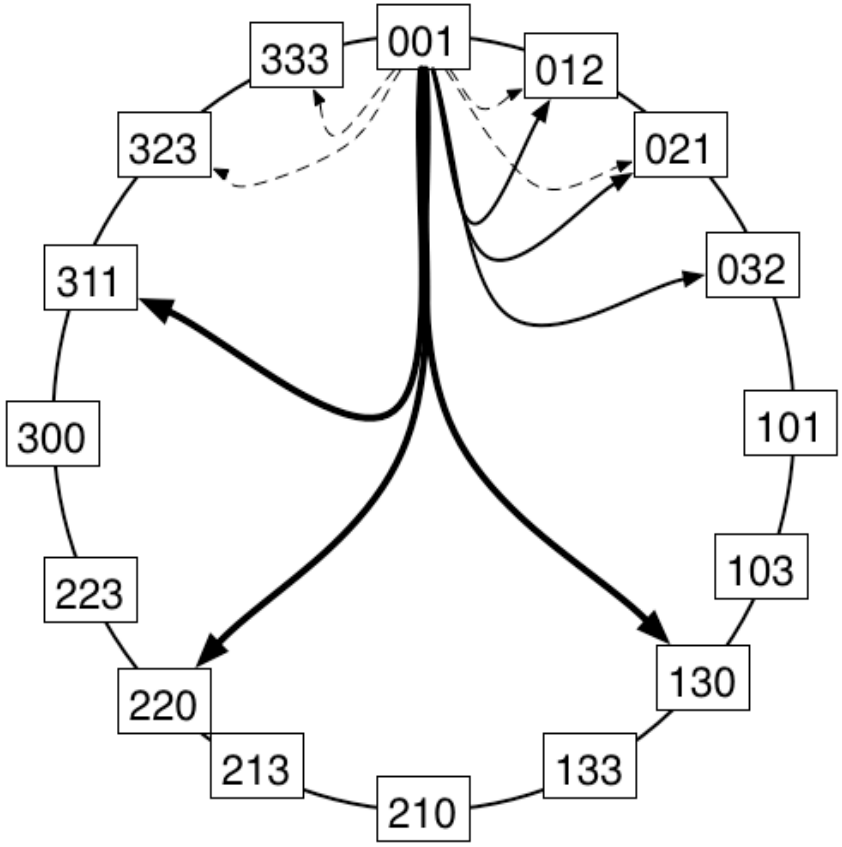
Suche in Chord

- **Theorem**
  - Die Suche in Chord benötigt O(log n) Hops mit hoher Wahrschein-lichkeit



Pastry

- Peter Druschel
  - jetzt Direktor des Max-Planck-Instituts für Informatik, Saarbrücken/ Kaiserslautern
- Antony Rowstron
  - Microsoft Research, Cambridge, GB
- Pastry
  - *Scalable, decentralized object location and routing for large scale peer-to-peer-network*
  - Chord-ähnliches Netzwerk, welches das Routing von Plaxton, Rajamaran, Richa (1997) verwendet



CoNe Freiburg

Der Wettlauf nach dem konstanten Grad

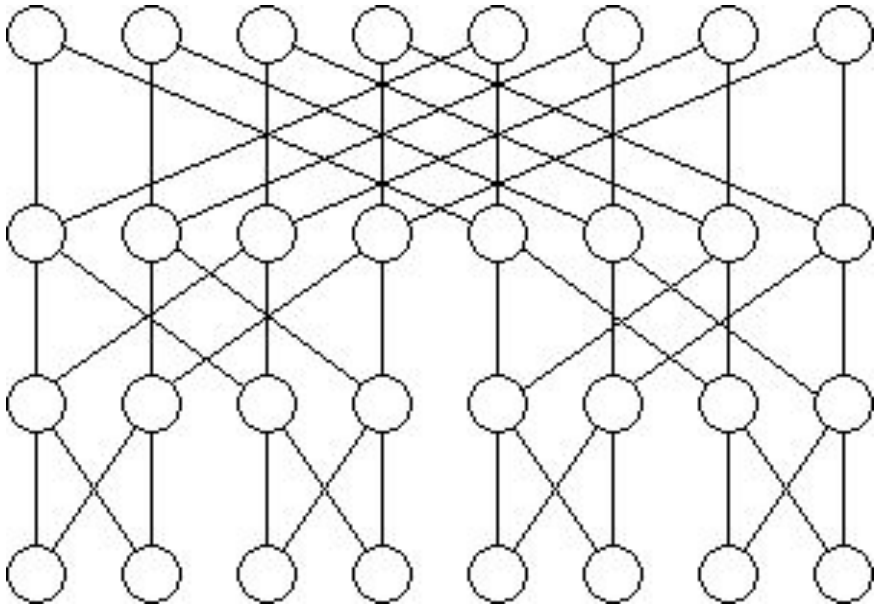
- CHORD:
  - Durchmesser O(log n)
  - Grad O(log n)
- Gesucht:
  - Netzwerk mit kleinem Ausgrad
  - D.h. Eingrad, Ausgrad konstant
  - Durchmesser O(log n)
- Lösung(en)
  - Viceroy
  - Koorde
  - Distance-Halving-Netzwerk

UNI FREIBURG

CoNe Freiburg

Viceroy 2001

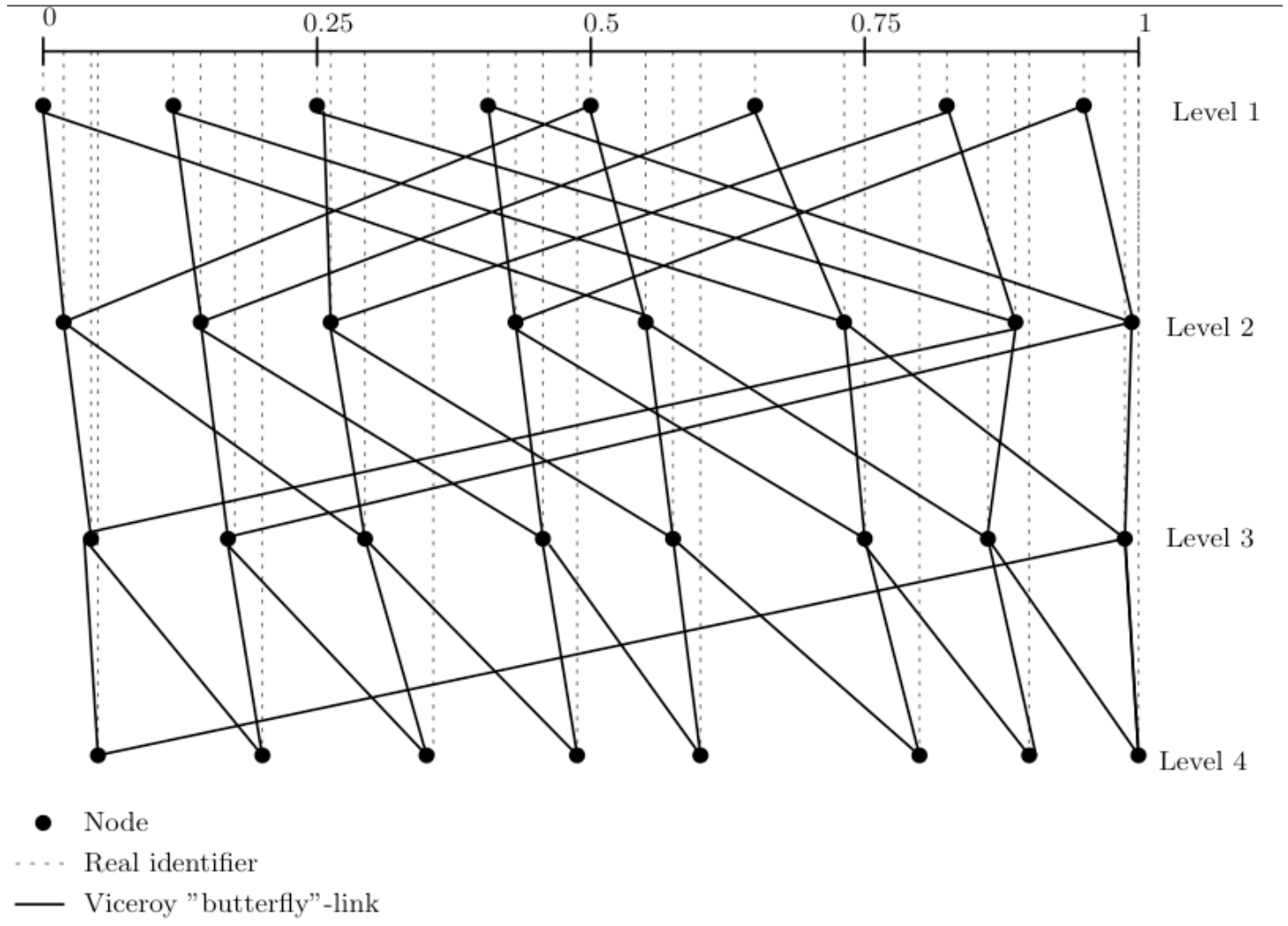
- Viceroy
  - A Scalable and Dynamic Emulation of the Butterfly
  - von Dahlia Malkhi, Moni Naor, David Ratajczak, 2001
- verwendet Butterfly-Graph



UNI FREIBURG

CoNe Freiburg

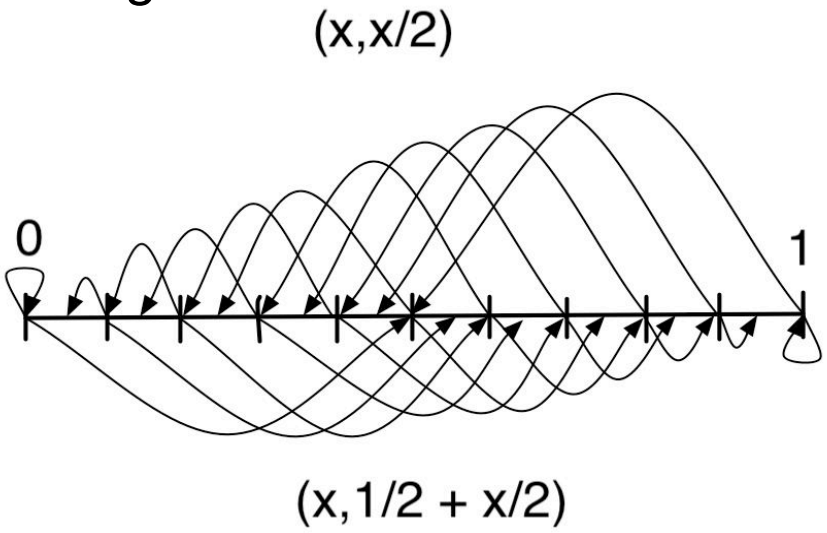
Aufbau Viceroy



CoNe Freiburg

Distance Halving

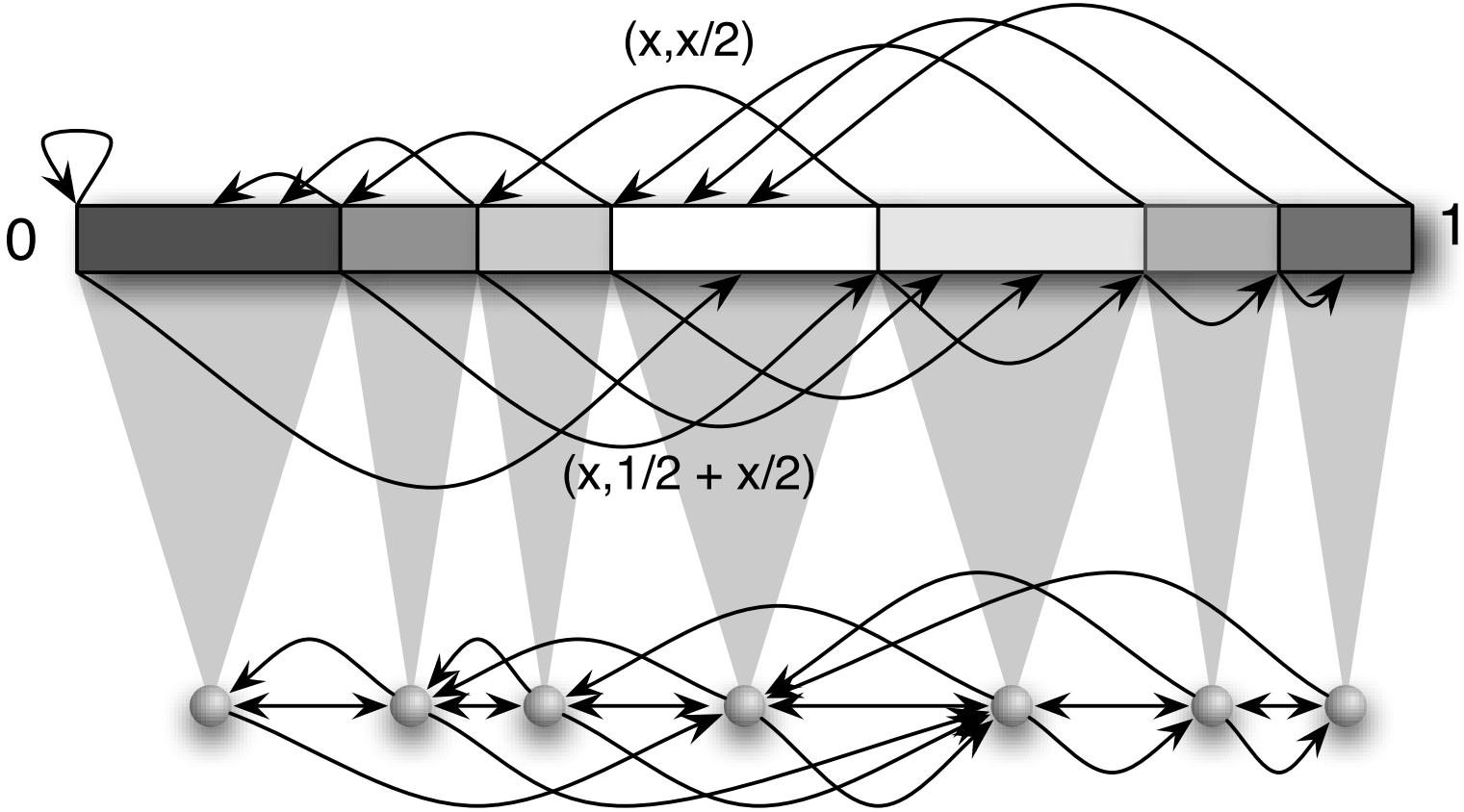
- Distance Halving von
  - Moni Naor und Udi Wieder, 2003
- Kontinuierliche Graphen
  - Sind unendliche Graphen mit kontinuierlicher Knotenmenge und Kantenmenge
- Der verwendete Graph
  - Knoten: $x \in [0,1)$
  - Bidirektionale Kanten
    - Links-Kanten (x,x/2)
    - Rechts-Kanten (x,1+x/2)



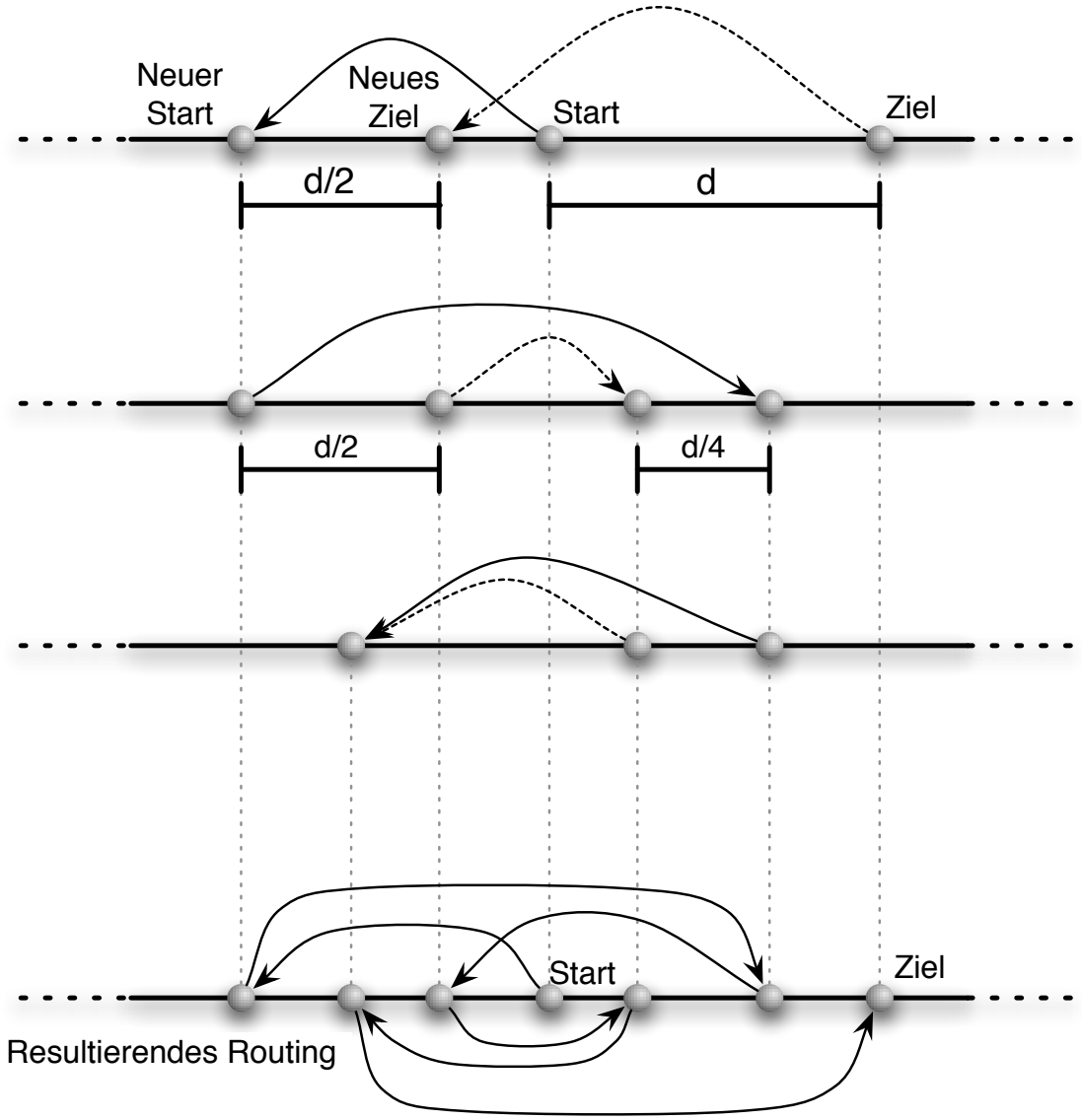
UNI FREIBURG

CoNe Freiburg

Abbildung von Peers auf Kantenbereiche



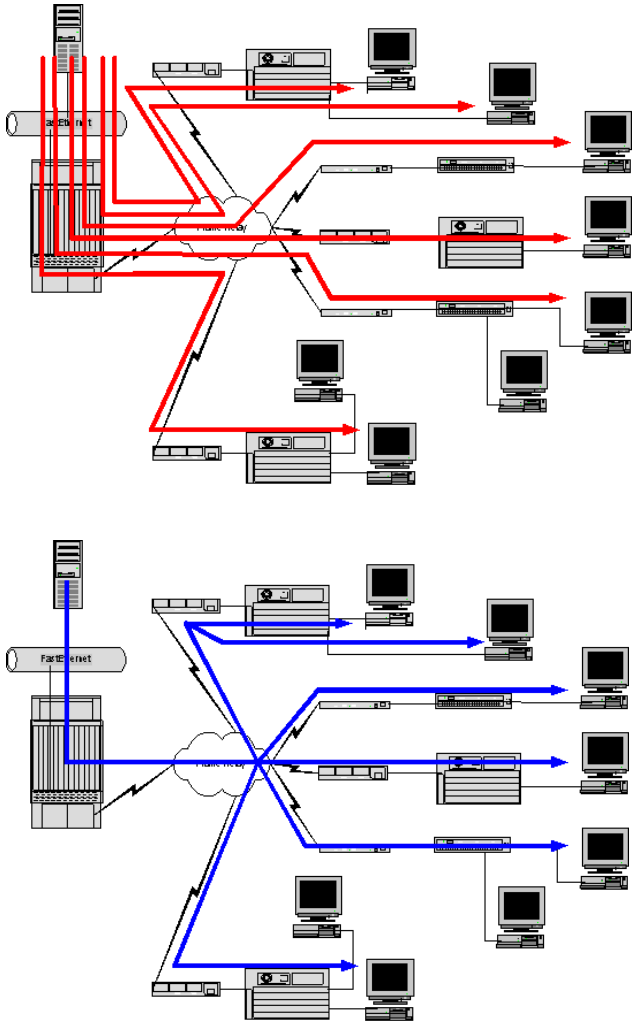
Logarithmische Suche in Distance-Halving



CoNe Freiburg

Der schnelle Download

- Problem
  - Wie kann eine Datei an ganz viele Teilnehmer verteilt werden?
- Unicast
  - Viele Unicast-Verbindungen (TCP) überlasten Server-Verbindung
- IP Multicast
  - Replikation der Pakete an den IP Routers

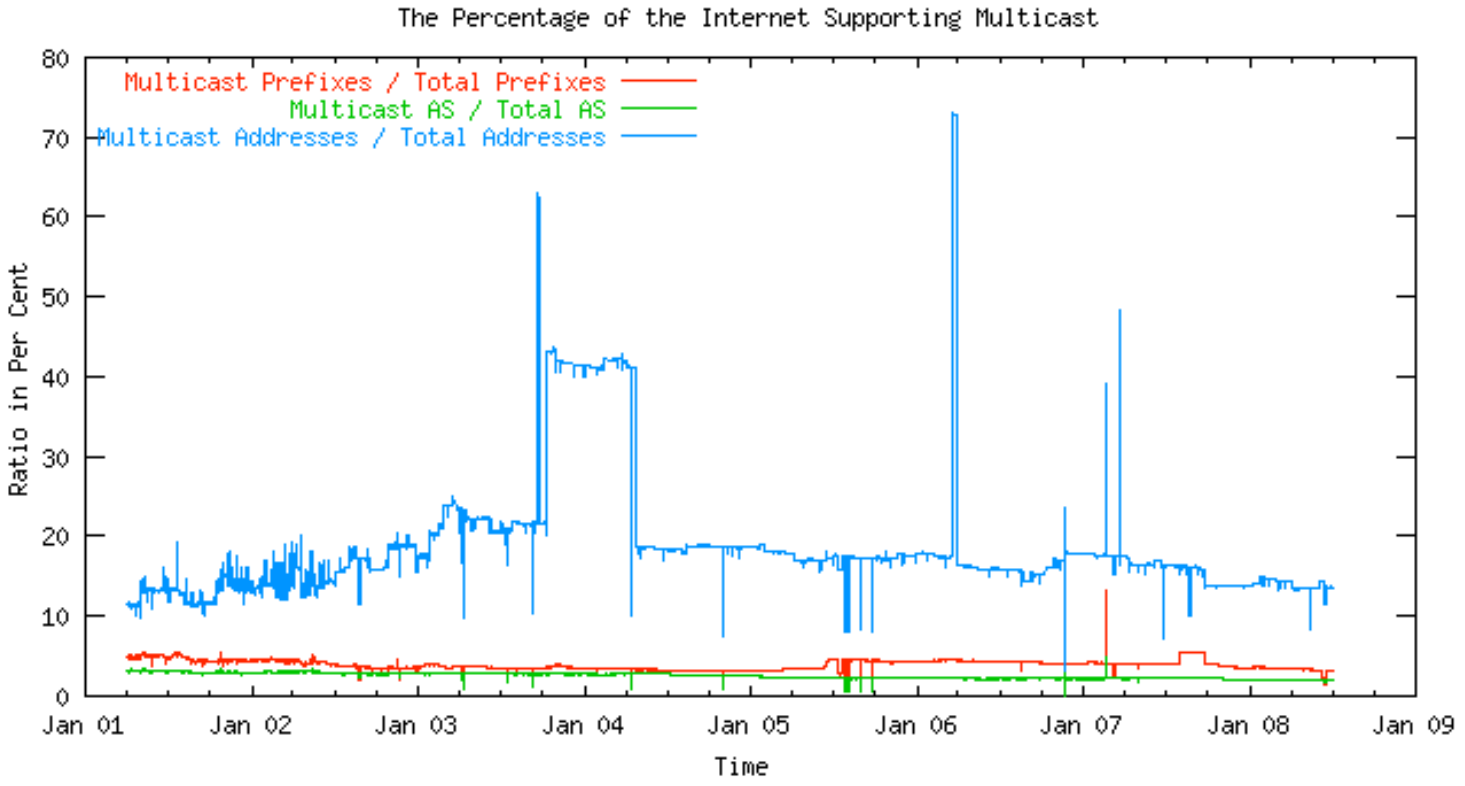


UNI FREIBURG

CoNe Freiburg

IP Multicast ist aber kaum verfügbar

- IP Multicast ist die schnellste Lösung für Dateiverteilung
- Weniger als 5% aller IP Router gestatten IP Multicast
  - http://www.multicasttech.com/status/

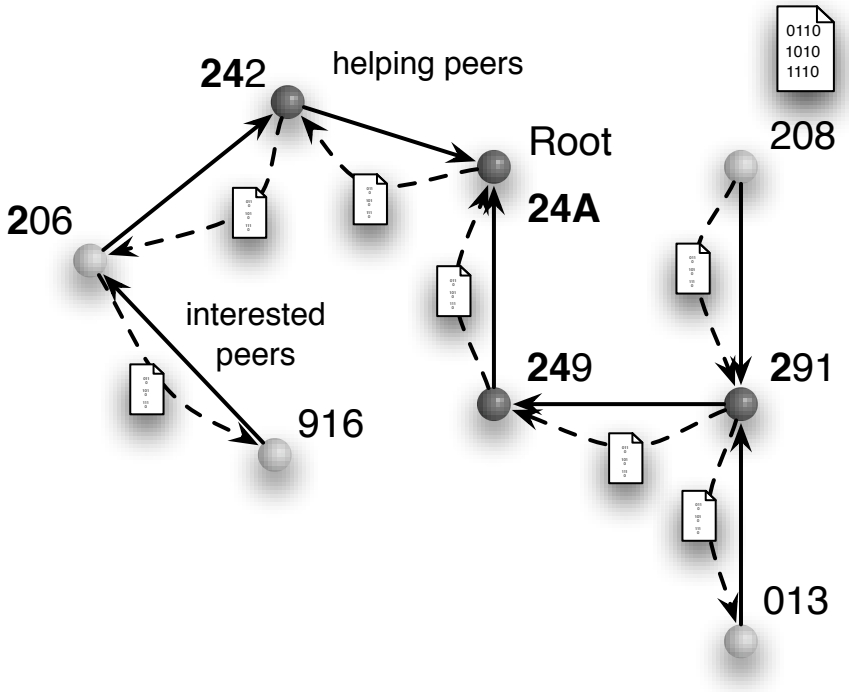


UNI FREIBURG

CoNe Freiburg

Multicast in P2P

- Multicast-Tree im Overlay (P2P) Network
- Scribe [2001] verwendet Pastry
  - Castro, Druschel, Kermarrec, Rowstron
- Ähnliche Ansätze
  - CAN Multicast [2001] based on CAN
  - Bayeux [2001] based on Tapestry

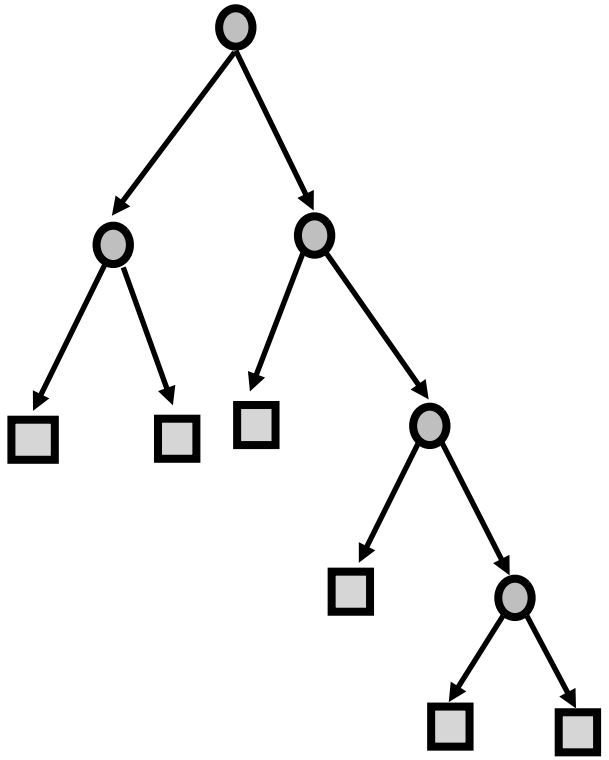


UNI FREIBURG

CoNe Freiburg

Split-Stream Motivation

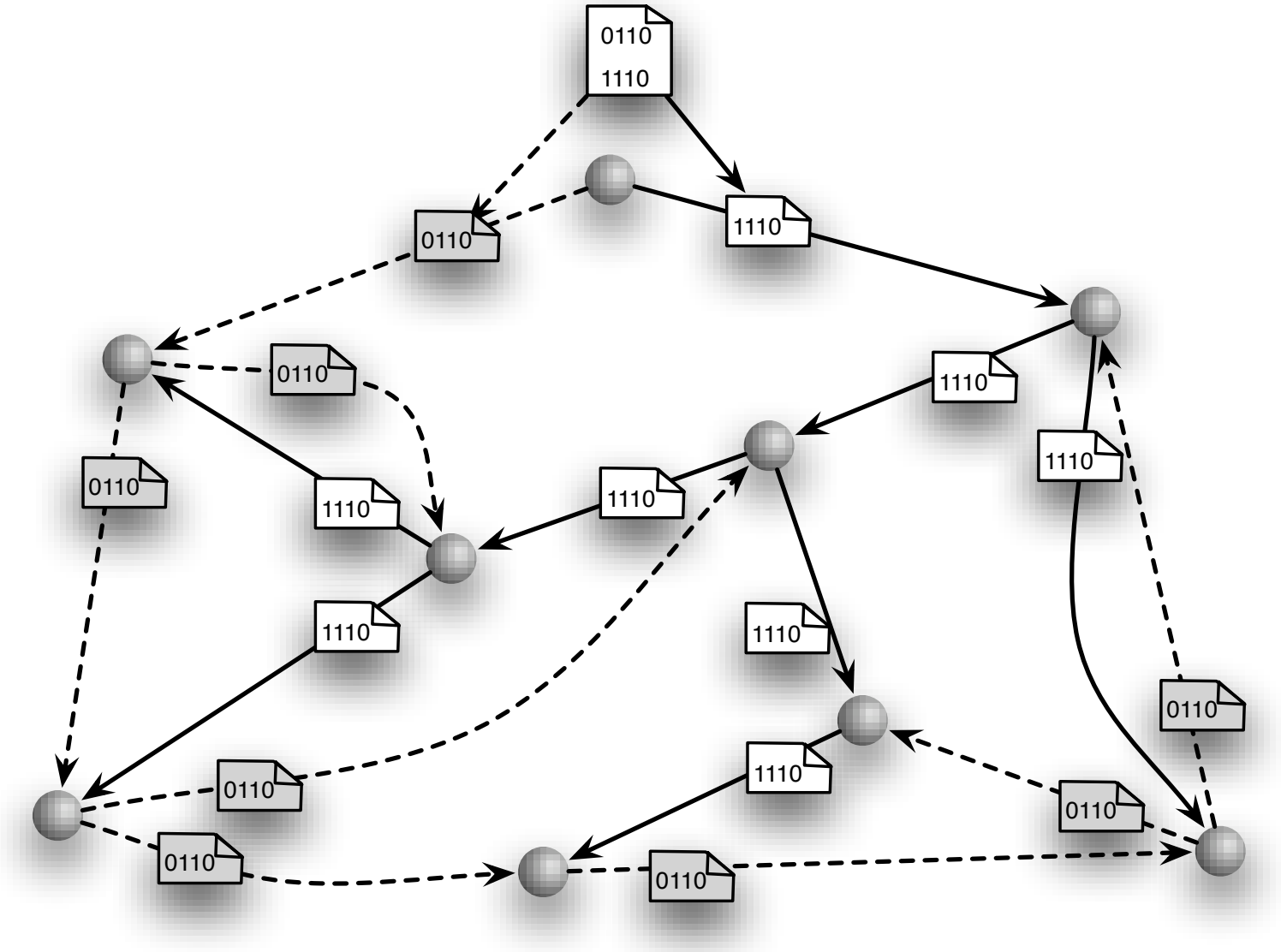
- Bäume diskriminieren Blätter-Peers
- Lemma
  - Anzahl der Blätter im Baum ist größer als die Anzahl der internen Knoten



UNI FREIBURG

CoNe Freiburg

Split-Stream



CoNe Freiburg

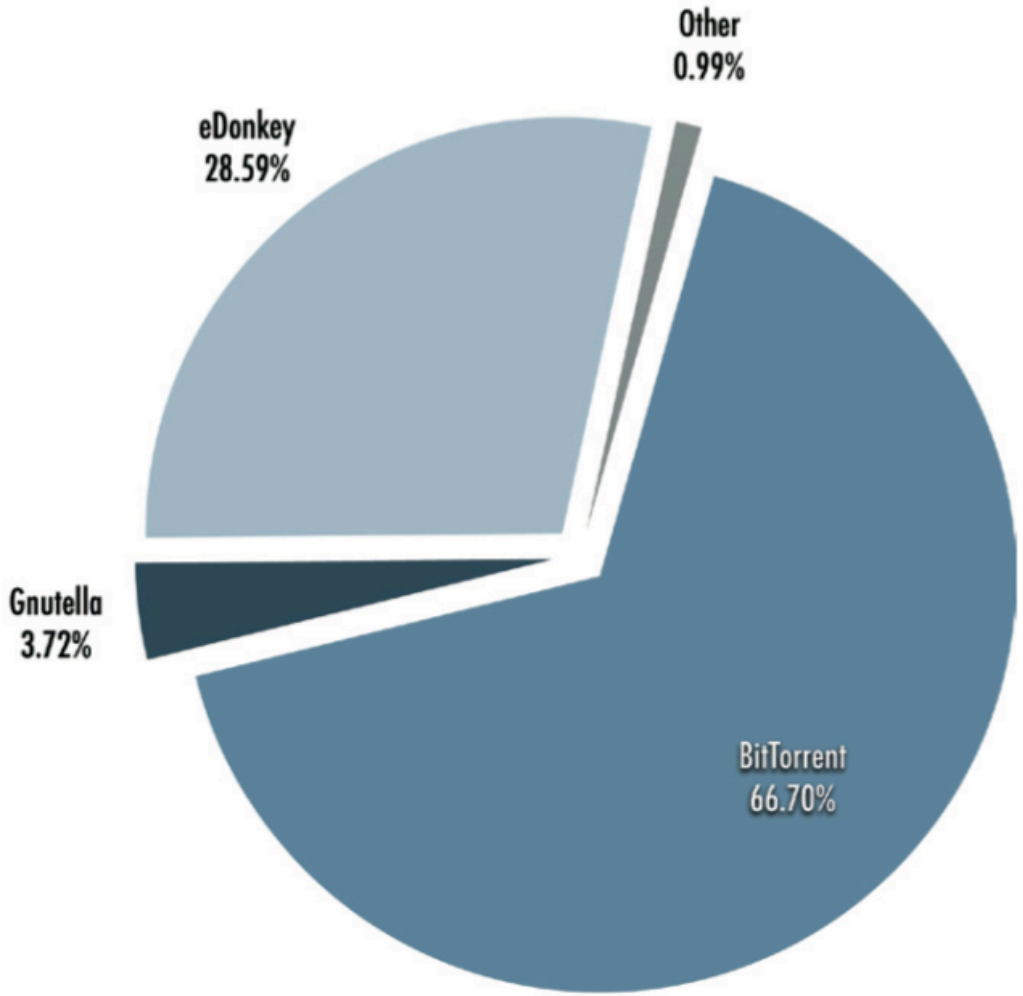
BitTorrent

- Bram Cohen
  - BitTorrent ist ein P2P-Netzwerk für den Download von Dateien
  - Dateien werden in Blöcke aufgeteilt
  - verwendet implizit Multicast-Bäume für die Verteilung von Blöcken
- Ziele
  - schneller Download einer Datei unter Verwendung des Uploads vieler Peers
    - Upload ist der Flaschenhals
    - z.B. wegen asymmetrischen Aufbau von ISDN oder DSL
  - Fairness
    - seeders against leeches
  - Gleichzeitige Verwendung vieler Peers

UNI FREIBURG

CoNe Freiburg

P2P Systems Germany 2007 by Volume



Quelle: Ipoque 2007

CoNe Freiburg

Probleme in Peer-to-Peer-Netzwerke

- Gnutella-Studie von Adar & Huberman 2000
  - ~70% der peers bieten keine Dateien an (free-riders)
  - Top 1% bieten 37% aller Dateien an
  - Ähnliches wird in Napster beobachtet

- 2005: 85% der Gnutella-Peers sind Free-Riders

UNI FREIBURG

CoNe Freiburg

Warum sind P2P-Benutzer egoistisch?

- Psyche der Benutzer
- Keine zentrale Autorität
- Dynamisches Nutzerverhalten
- Verfügbarkeit günstiger Identitäten
- Versteckte und nicht verfolgbare Aktionen
- Absichtlicher Betrug

- **Folgerung**
  - Peer-to-Peer-Netzwerke müssen mit egoistischen Nutzern fertig werden

UNI FREIBURG

CoNe Freiburg

Gefangenen-Dilemma

- Prisoner‘s dilemma (Flood&Drescher 1950)
  - zwei Verdächtige
  - wenn einer gesteht und der andere schweigt, dann kriegt der Kollaborateur keine Strafe und der andere 10 Jahre
  - wenn beide gestehen bekommen Sie jeweils 7 Jahre
  - wenn keiner gesteht, dann bekommen sie ein Jahr Gefängnis
- Beste soziale Strategie
  - beide schweigen
- Nash equilibrium
  - Wenn jede Partei unter der Annahme, dass die anderen konstant bleiben seine Kosten optimiert
  - Hier: Beide reden

	A redet	A schweigt
B redet	A: -7 B: -7	A: -10 B: 0
B schweigt	A: 0 B: -10	A: -1 B: -1

UNI FREIBURG

Gefangenen-Dilemma des Peer-to-Peer Filesharing

	U: Peer uploads	U: Peer rejects upload
D: Peer downloads	D: 10 U: -1	D: 0 U: 0
D: Peer does not download	D: 0 U: 0	D: 0 U: 0

CoNe Freiburg

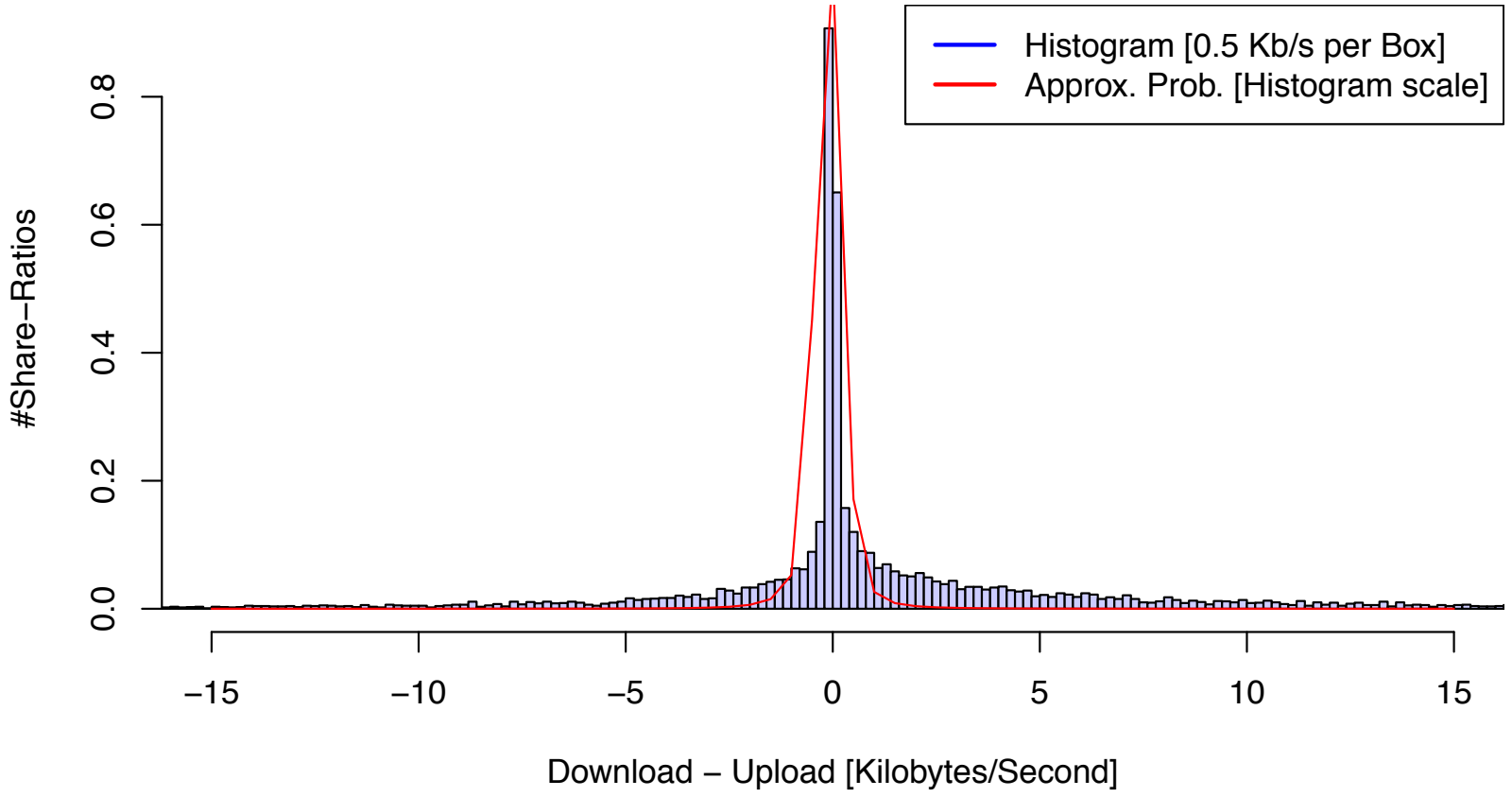
Bittorrent – Policy

- Ziel
  - selbst organisierendes System
  - gute Peers (hoher Upload, geringer Download) werden belohnt
  - schlechte Peers (hoher Download, kaum Upload) werden bestraft
- Belohnung
  - gute Download-Geschwindigkeit
  - „un-choking“
- Bestrafung
  - „choking“ (Drosseln) der Download-Geschwindigkeit
- Bewertung
  - Jeder Peer bewertet seine Umgebung ausgehend von seiner persönlichen Erfahrung

UNI FREIBURG

CoNe Freiburg

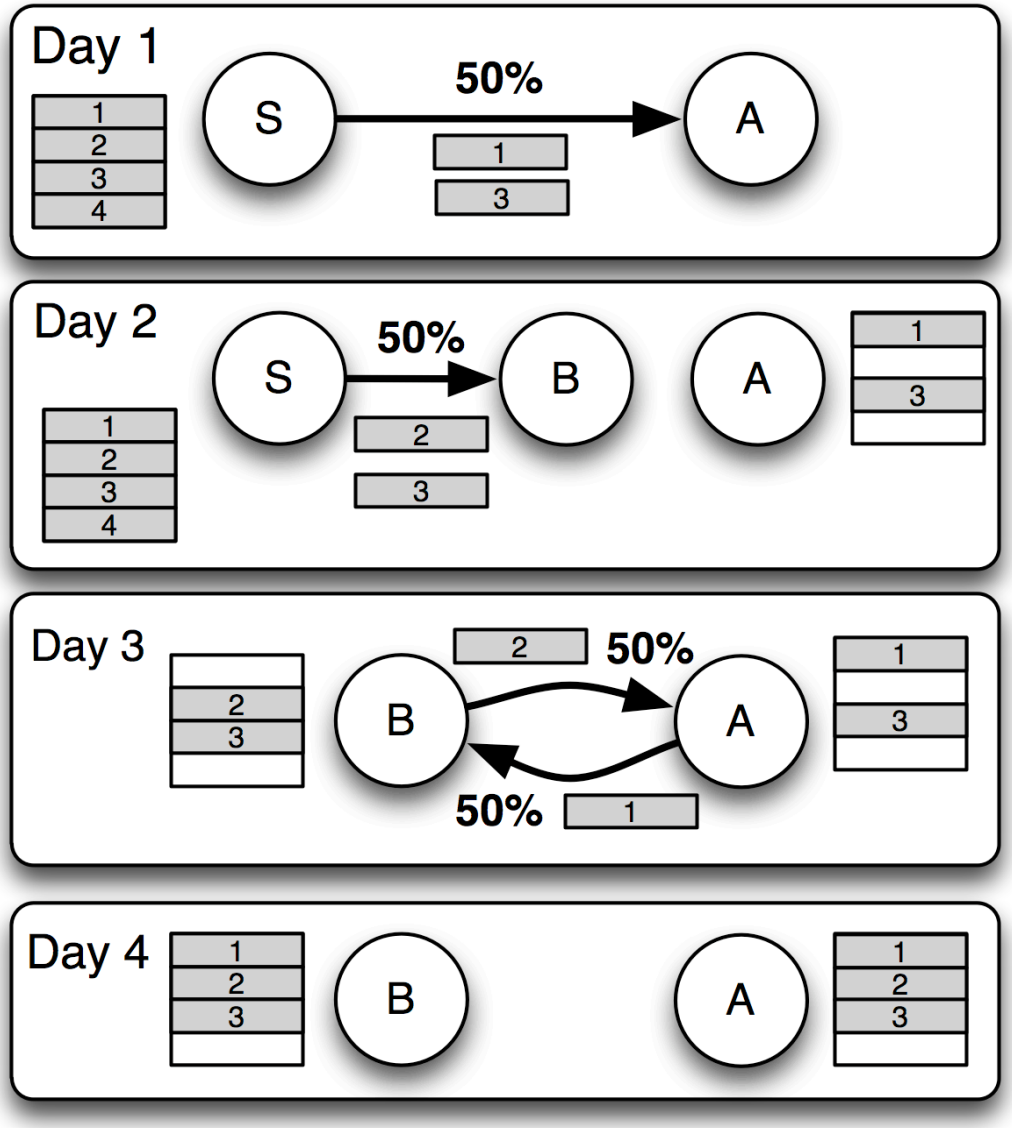
Fairness BitTorrent August 2009



UNI FREIBURG

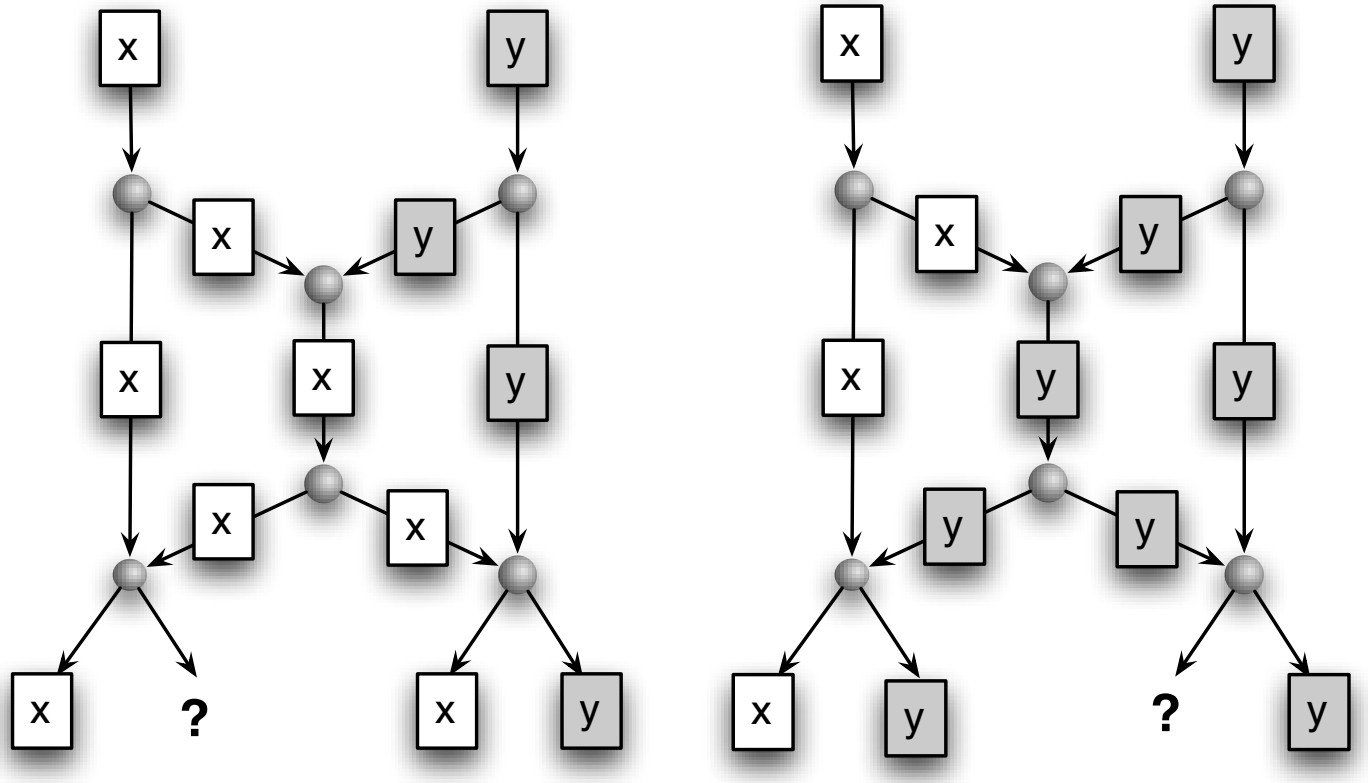
CoNe Freiburg

Probleme mit BitTorrent



CoNe Freiburg

Network Coding



CoNe Freiburg

Network Coding

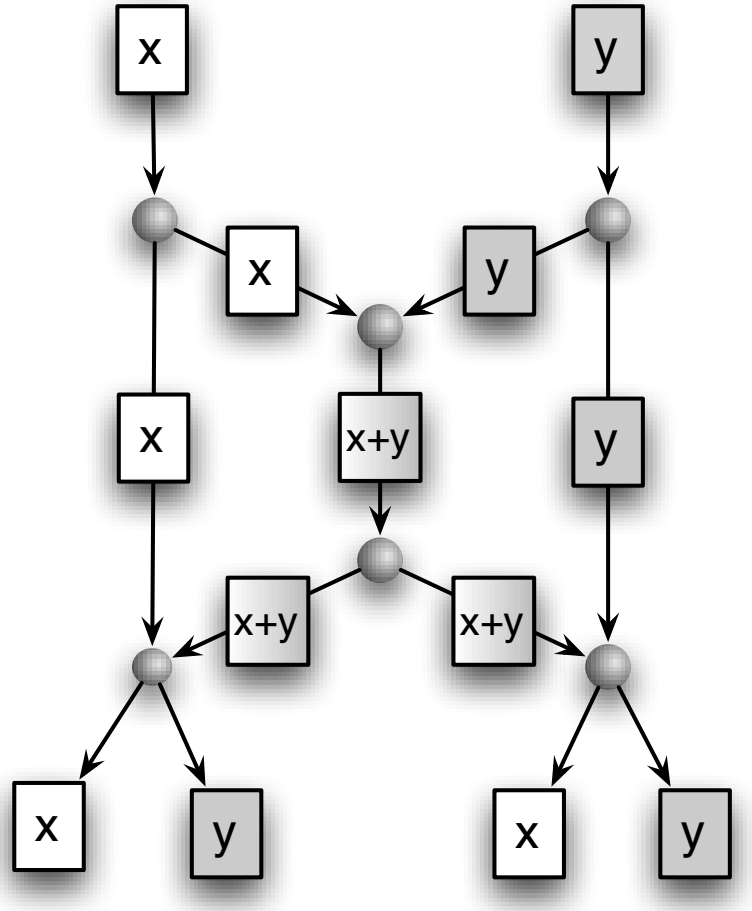
- R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network Information Flow", (IEEE Transactions on Information Theory, IT-46, pp. 1204-1216, 2000)

- Theorem [Ahlswede et al.]
  - Es gibt für jeden Graphen einen Netzwerk-Code, so dass jeder Knoten so viel Information bekommt wie es der maximale Fluss von den Quellen zu diesen Knoten ermöglicht.

UNI FREIBURG

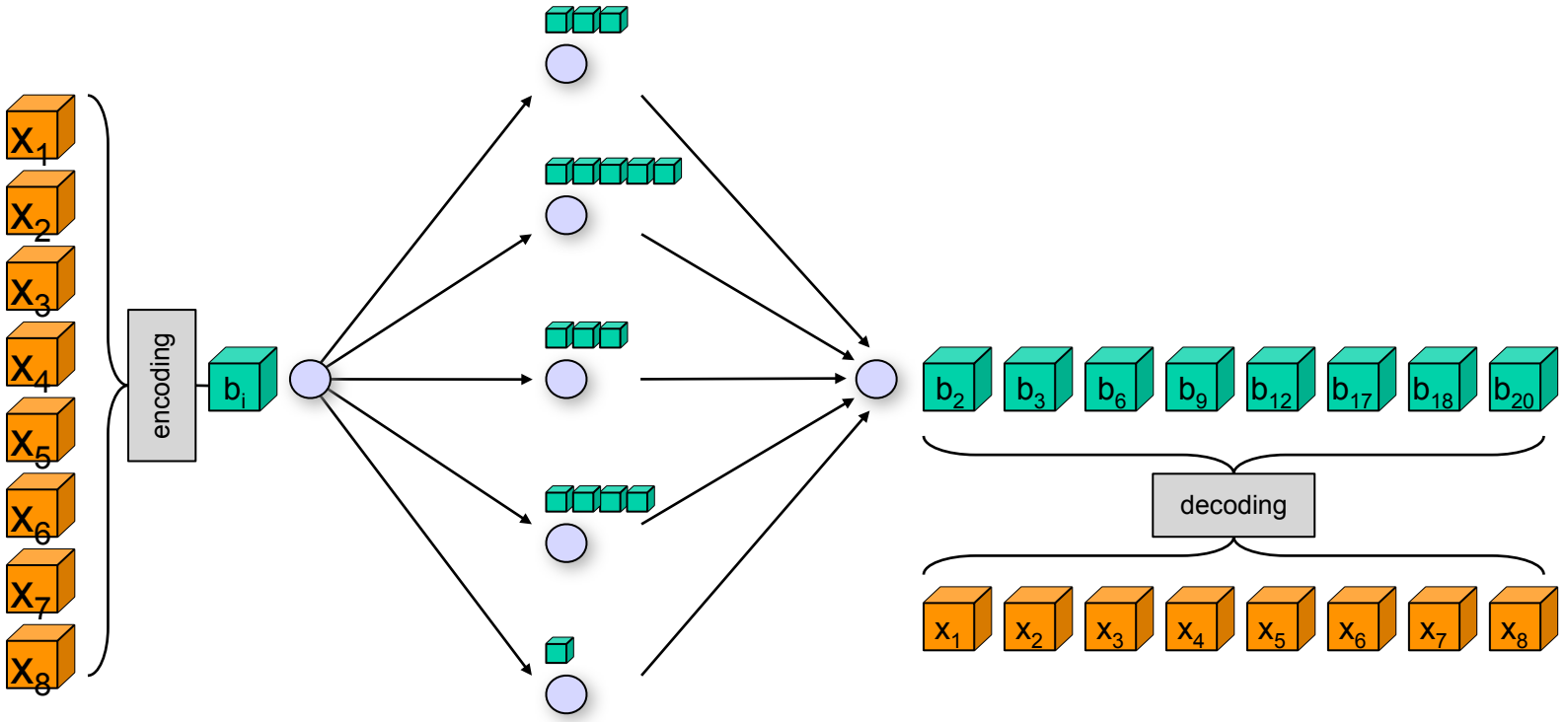
CoNe Freiburg

Network Coding



CoNe Freiburg

Kodierung und Dekodierung



CoNe Freiburg

Lineare Netzwerk-Codes

- Datei X = (x_1, x_2, ..., x_n)
- Koeffizienten c_{ij}
- Code-Blocks b_1, b_2, ..., b_n

$$(c_{i1}, \ldots, c_{in}) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = b_i$$

$$\begin{pmatrix} c_{11} & \ldots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \ldots & c_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

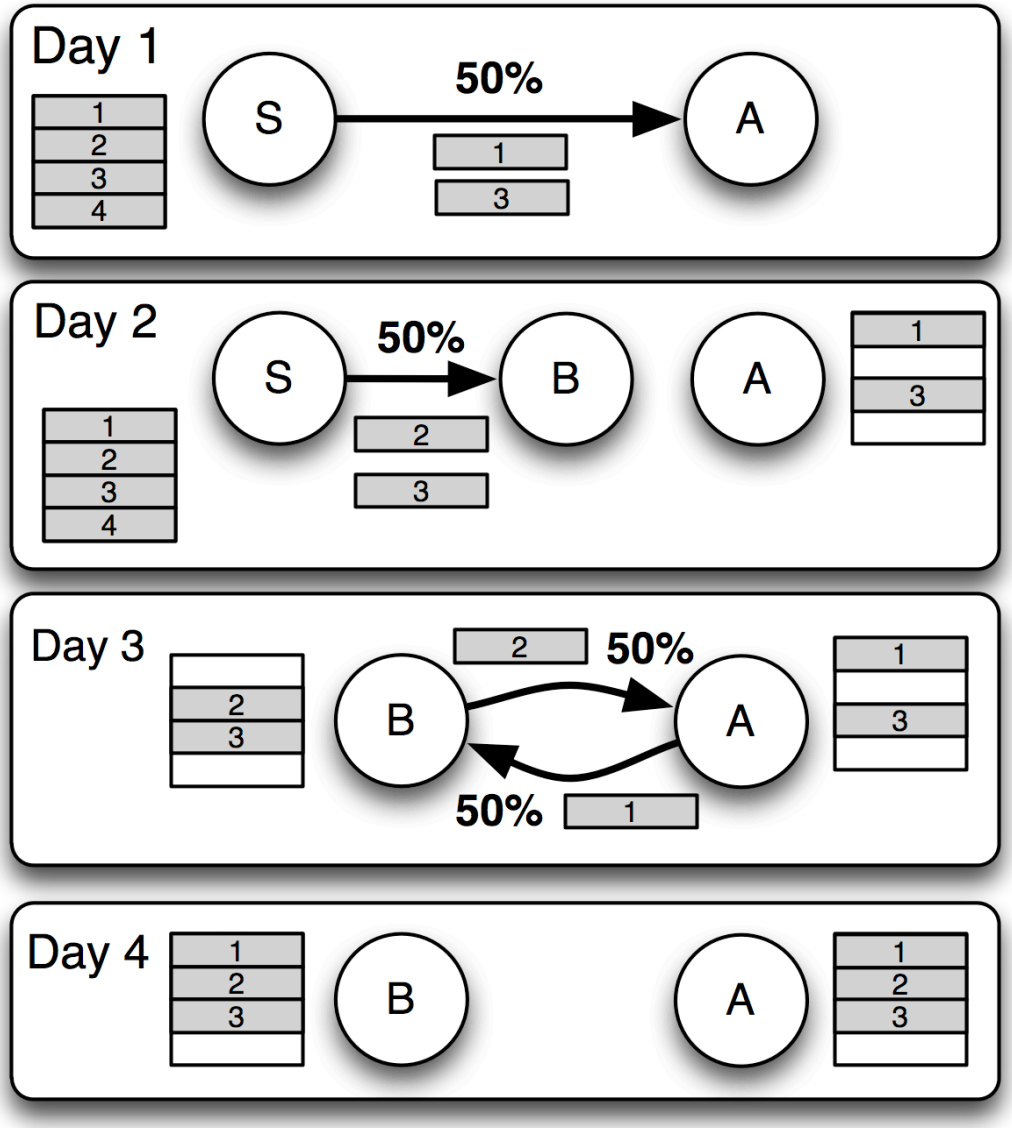
- Falls die Matrix invertierbar ist:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_{11} & \ldots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \ldots & c_{nn} \end{pmatrix}^{-1} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

UNI FREIBURG

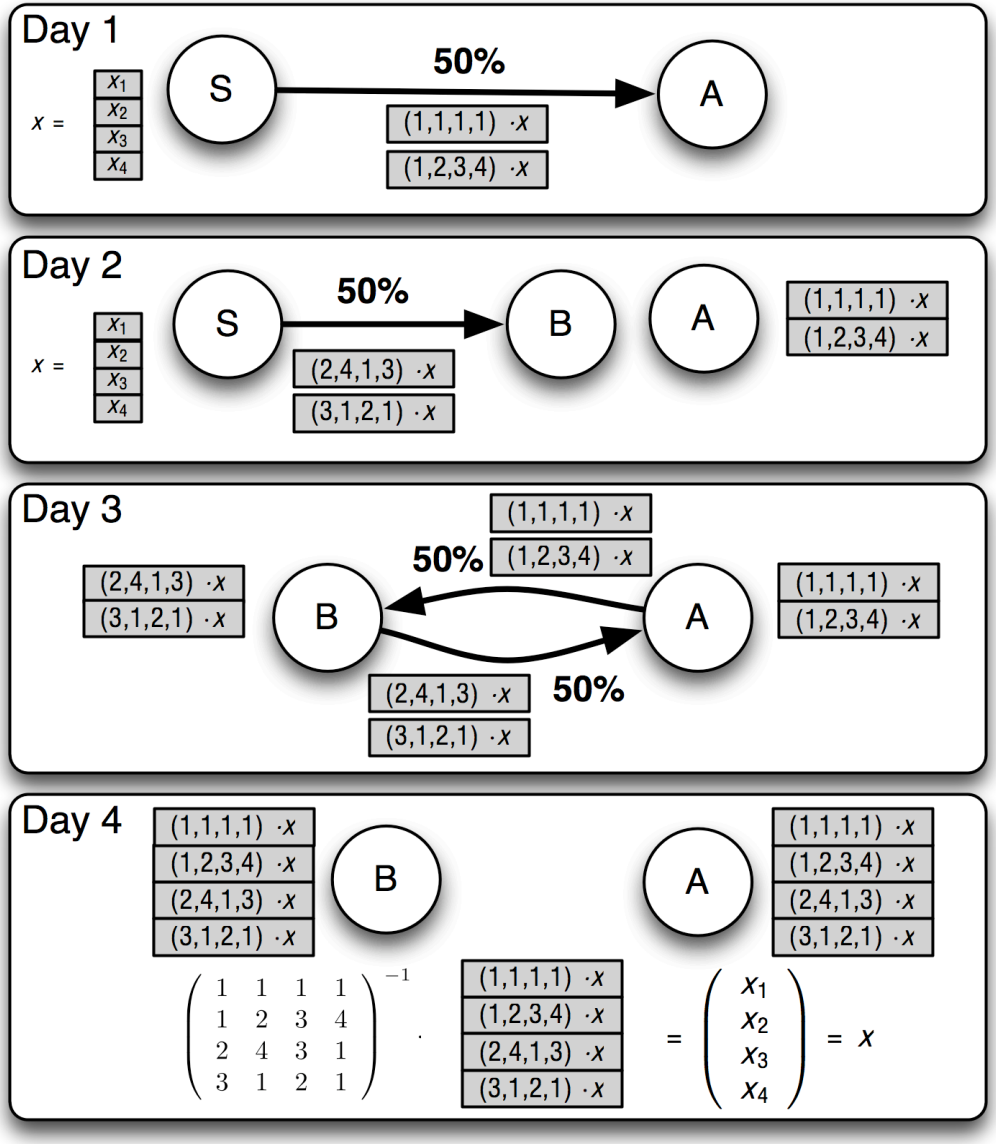
CoNe Freiburg

Probleme mit BitTorrent



CoNe Freiburg

Lösung durch Network Coding



CoNe Freiburg

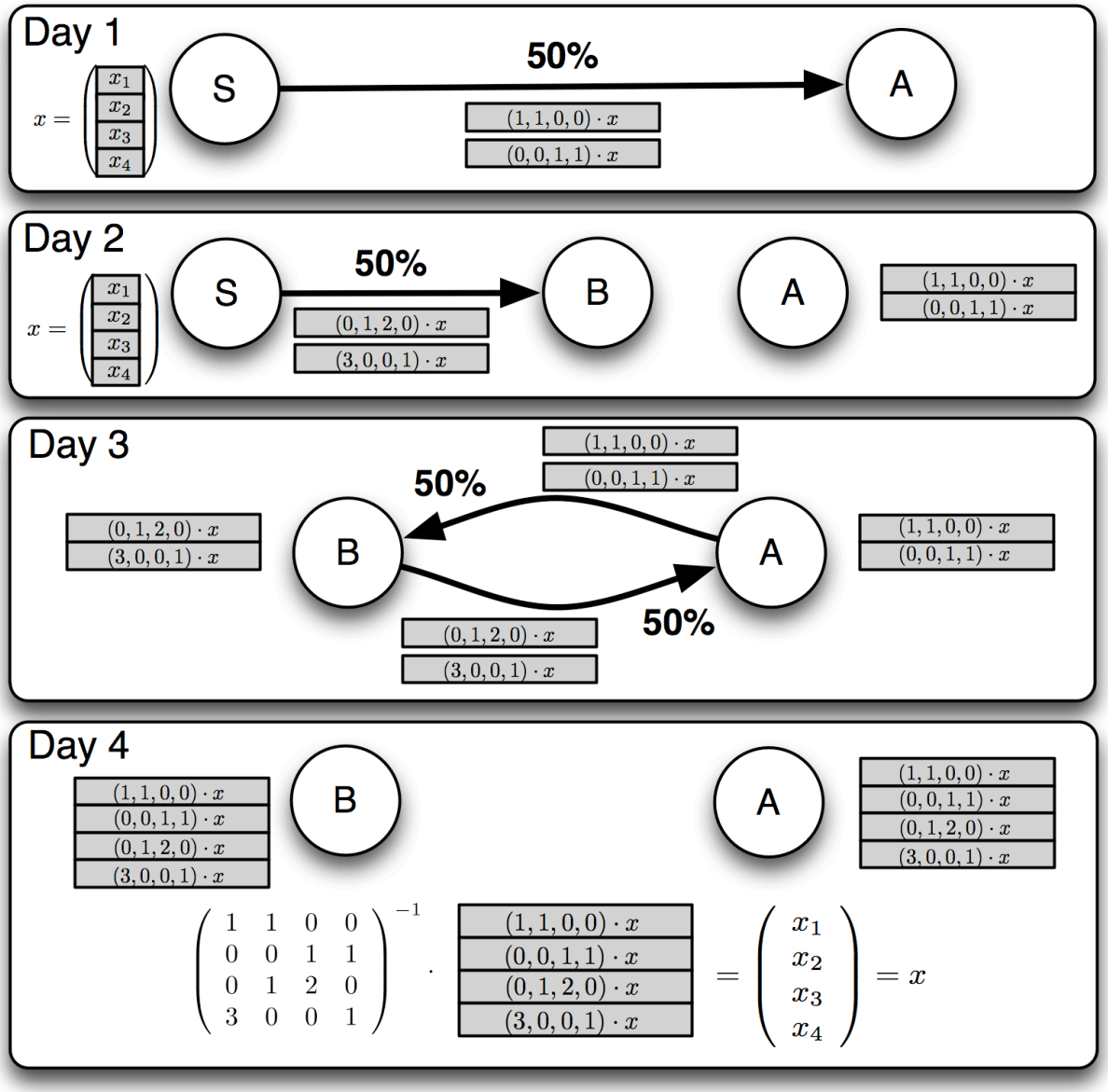
Probleme der Network Coding

- Lese/Schreib-Zugriffe auf die Festplatte
  - Das Dekodieren oder Kodieren eines Code-Blocks benötigt lineare Zeit
  - Zum Schreiben von 100 Blöcken eines 4 GB großen Datei müssen 400 GB von der Festplatte gelesene werden!

UNI FREIBURG

CoNe Freiburg

Paircoding

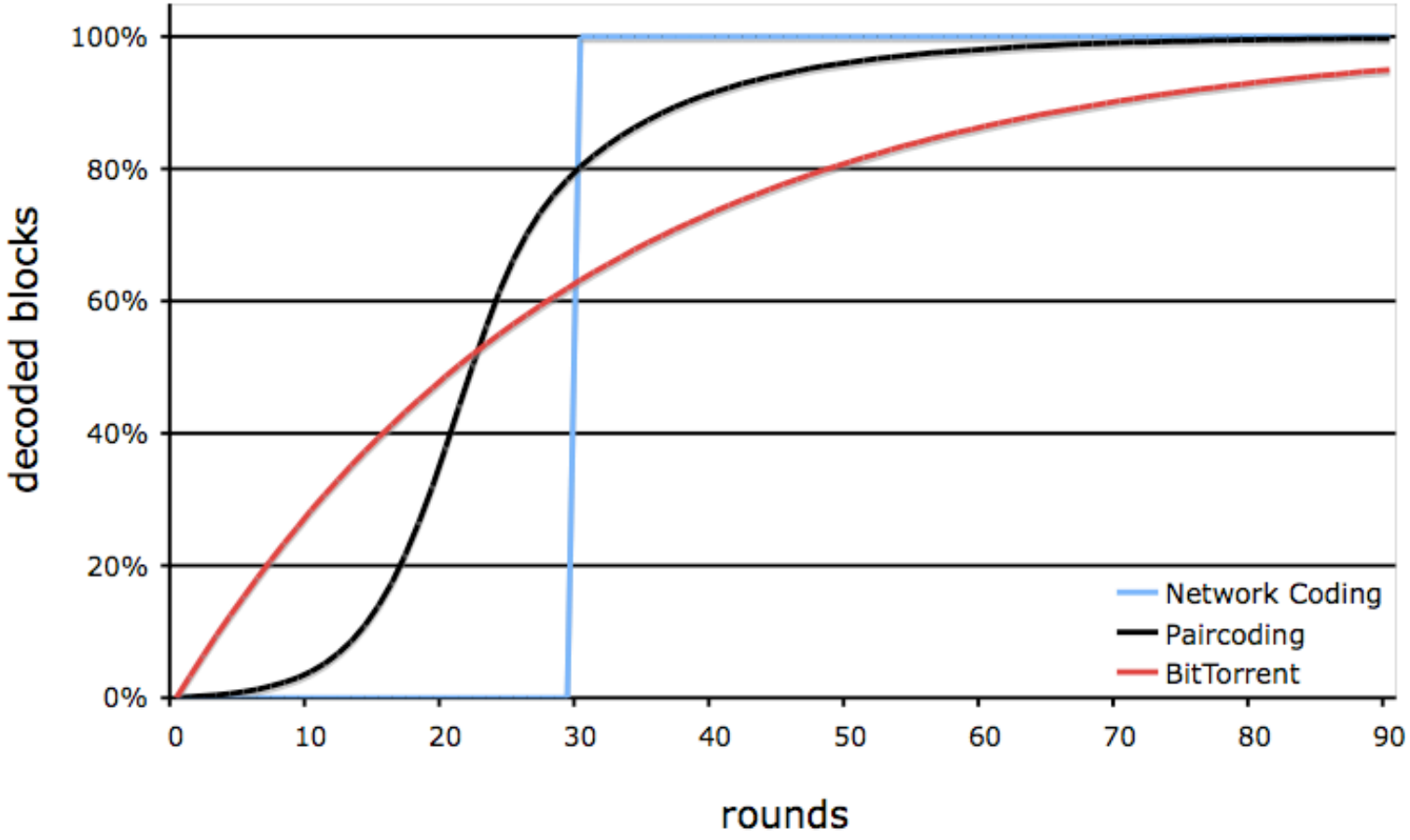


UNI FREIBURG

CoNe Freiburg

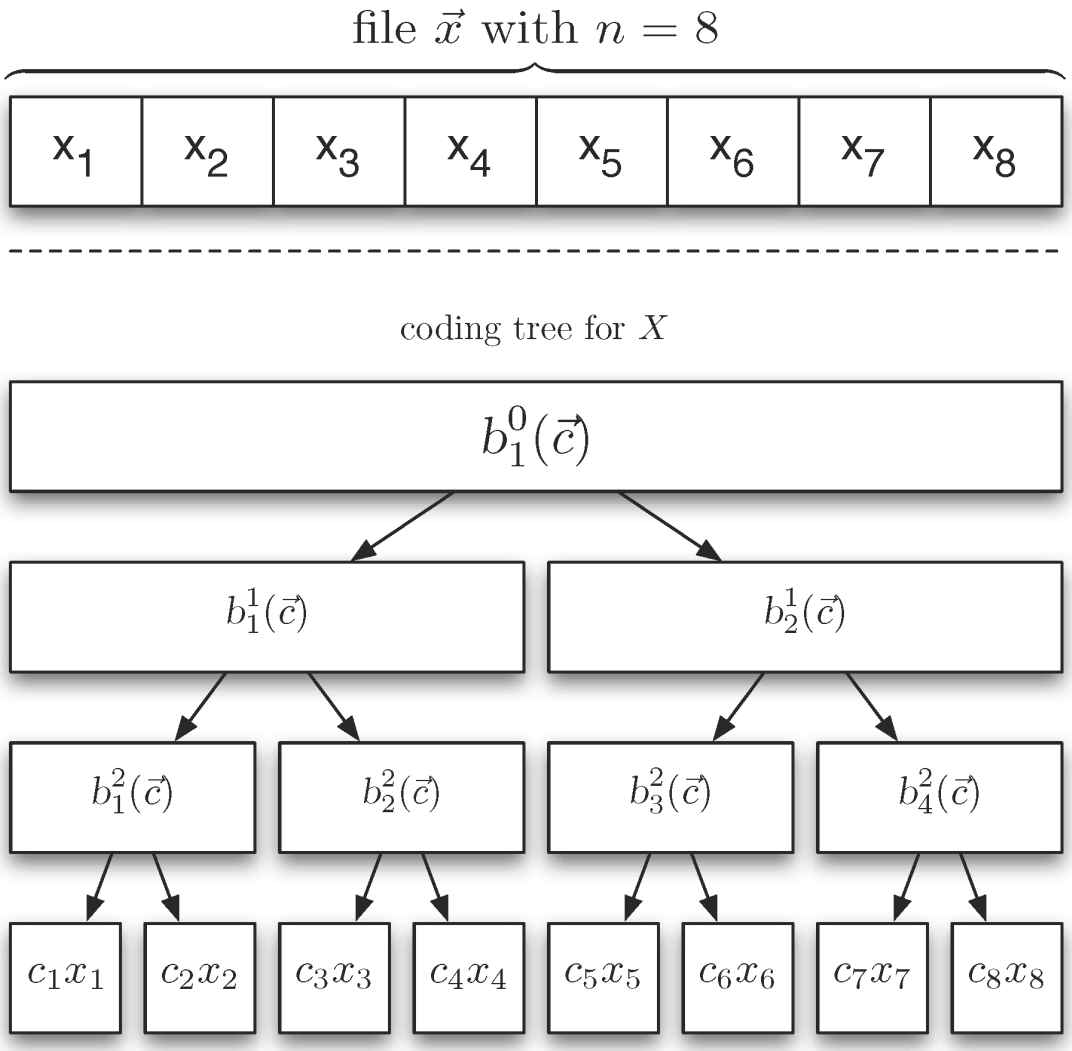
Vergleich Netzwerk-Codierung, Pair-Coding und BitTorrent

- Seed sendet zufällige Blöcke in jeder Runde



CoNe Freiburg

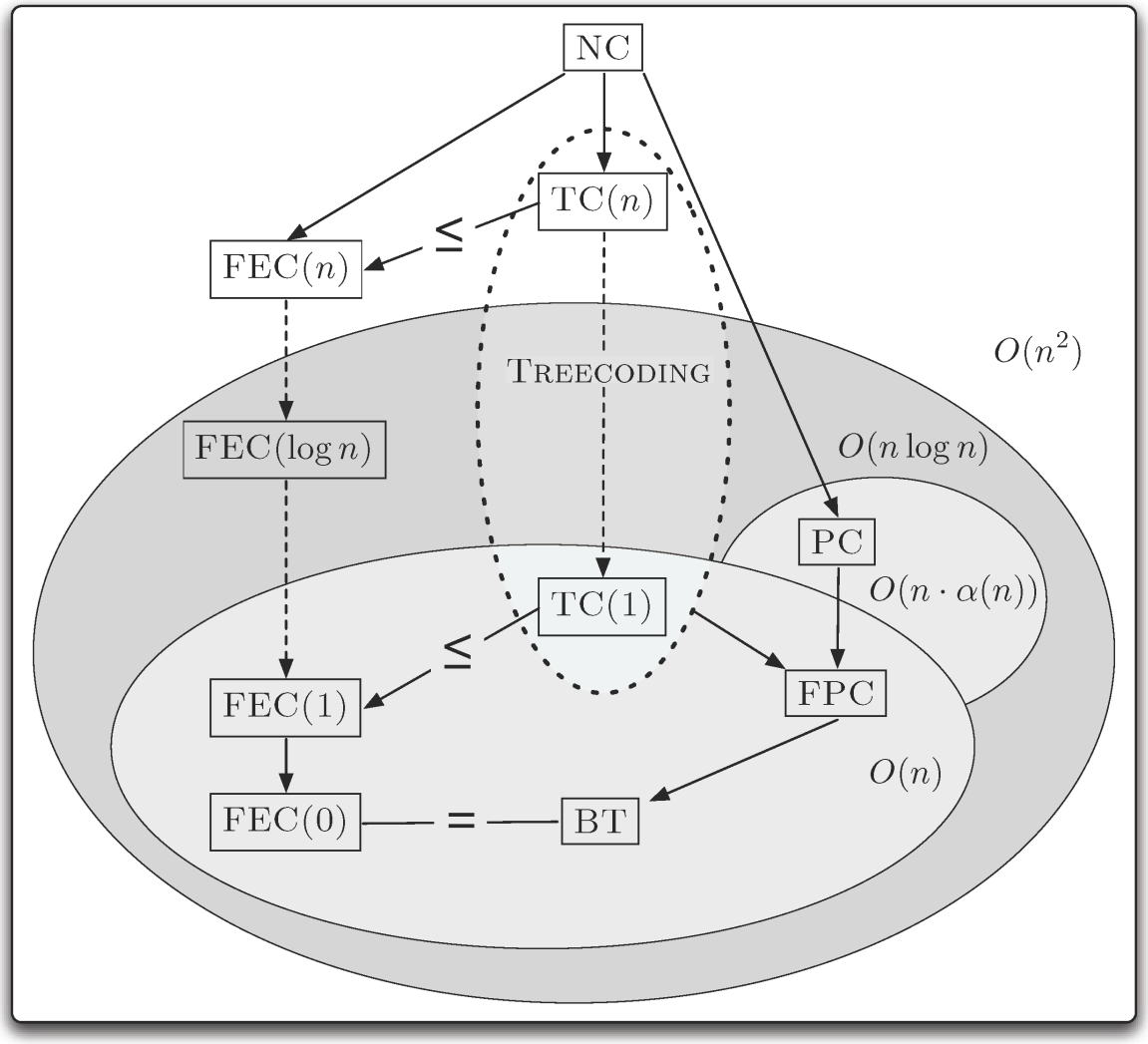
Noch bessere Codes: Tree-Coding



UNI FREIBURG

CoNe Freiburg

Klassifizierung von Netzwerk-Codes



CoNe Freiburg

Zusammenfassung und Ausblick

- Peer-to-Peer-Netzwerke
  - Hoch-Phase 2007
  - sind erwachsen geworden
  - haben sich etabliert
- Theorie und Praxis
  - Algorithmen wie DHT werden jetzt eingesetzt
    - z.B. verteilte Tracker für BitTorrent
- P2P und das Gesetz
  - Verfolgung des Filesharings verändert die P2P-Netzwerke
  - Zentrale Strukturen werden abgeschafft
  - Anonymisierung wird attraktiver
  - Filesharing lässt sich auch langfristig nicht verhindern

UNI FREIBURG

CoNe Freiburg

IIF INSTITUT FÜR INFORMATIK FREIBURG

Algorithmen und Datenstrukturen für Peer-to-Peer-Netzwerke

Christian Schindelhauer
Technische Fakultät
Rechnernetze und Telematik
Albert-Ludwigs-Universität Freiburg

UNI FREIBURG