

Organic Computing Peer-to-Peer-Netzwerke

Rolf Wanka

Sommersemester 2008

26. Juni 2008

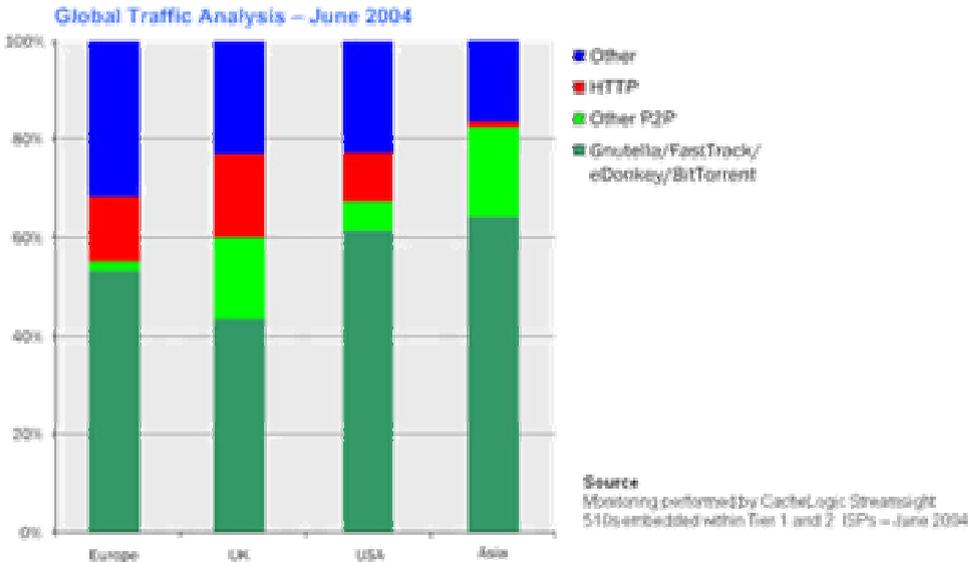
rwanka@cs.fau.de



P2P-Netzwerke aktuell

- Juni 2004

–Quelle: CacheLogic



- 2005

–Über 8 Mio. aktive Teilnehmer an Peer-to-Peer-Netzwerken zu jeder Zeit

–10 PetaByte an Daten zu jeder Zeit

–Mehr als die Hälfte des gesamten Internet-Traffic ist Peer-to-Peer

–Mehrere Peer-to-Peer-Netzwerke durch Gerichtsprozesse stillgelegt

–Tausende von Einzelklagen gegen Peer-to-Peer-Nutzer wegen Verletzung des Urnehmerschutzes

Meilensteine Praxis

- Napster (1999)
 - seit 1999, bis 2000 (Gerichtsurteil)
- Gnutella (2000)
 - Neue Version (Gnutella 2) in 2002
- Edonkey (2000)
 - Später: **Overnet** unter Verwendung von Kademlia
- FreeNet (2000)
 - Anonymisierung der Teilnehmer
- JXTA (2001)
 - Open Source Peer-to-Peer-Netzwerk-Plattform

Meilensteine Praxis

- FastTrack (2001)
 - bekannt durch KaZaa, Morpheus, Grokster
- Bittorrent (2001)
 - Nur Download-System, keine Suche
- ...



Meilensteine Theorie

- **Distributed Hash-Tables (DHT) (1997)**
 - Urspr. für Lastverteilung zwischen Web-Servern
- **CAN (2001)**
 - Effiziente verteilte DHT-Datenstruktur für P2P-Netzwerke
- **Chord (2001)**
 - Effiziente verteilte P2P-Datenstruktur mit logarithmischer Suchzeit
- **Pastry/Tapestry (2001)**
 - Effiziente verteilte P2P-Datenstruktur aufbauend auf Routing von Plaxton
- **Kademlia (2002)**
 - P2P-Lookup basierend auf XOR-Metrik
- **Viele weitere interessante Netzwerke**
 - **Viceroy**, Distance-Halving, Koorde, Skip-Net, P-Grid, ...



Was ist ein P2P-Netzwerk?

- Was ist ein Peer-to-Peer-Netzwerk nicht?
 - Ein Peer-to-Peer-Netzwerk ist kein Client-Server-Netzwerk!
- Ethymologie: Peer
 - heißt Gleicher, Ebenbürtiger, von lat. par
- Definition
 - *Peer-to-Peer*
 - bezeichnet eine Beziehung zwischen gleichwertigen Partnern
 - *P2P*
 - Internet-Slang für Peer-to-Peer
 - Ein *Peer-to-Peer-Netzwerk* ist ein
 - Kommunikationsnetzwerk zwischen Rechnern im Internet
 - in dem es keine zentrale Steuerung gibt
 - und keine zuverlässigen Partner.



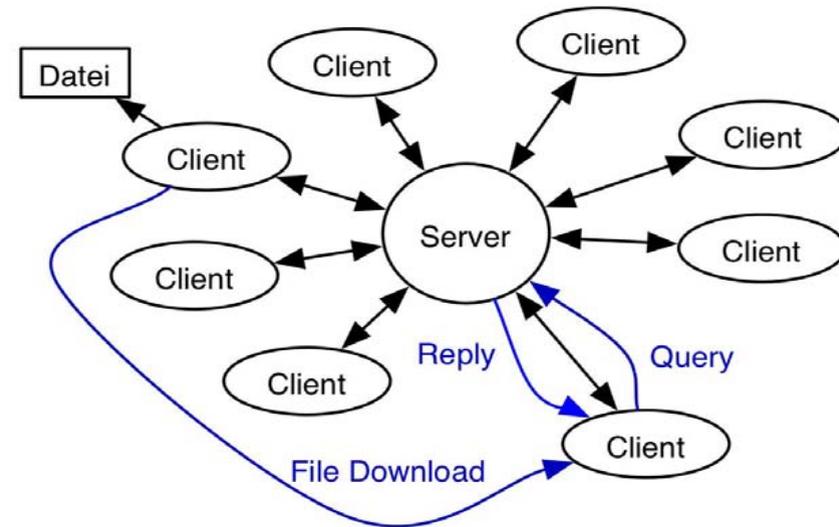
Was ist ein P2P-Netzwerk?

- Beobachtung
 - Das Internet ist (eigentlich auch) ein Peer-to-Peer-Netzwerk
 - Definitionen zu ungenau

- Shawn (Napster) Fanning
 - brachte Juni 1999 eine Beta-Version seines mittlerweile legendären Napster-Peer-to-Peer-Netzwerks heraus
 - Ziel: File-sharing-System
 - Tatsächlich: Musik-Tauschbörse
 - Herbst 1999 war Napster „Download des Jahres“
- Urheberrechtsklage der Musik-Industrie im Juni 2000
- Gegen Ende 2000 Kooperationsvertrag
 - zwischen Fanning und Bertelsmann Ecommerce
- Seitdem ist Napster eine kommerzielle File-Sharing-Plattform

Wie funktioniert Napster?

- Client-Server-Struktur
- Server unterhält
 - Index mit Meta-Daten
 - Dateiname, Datum, etc
 - Tabelle der Verbindungen der teilnehmenden Clients
 - Tabelle aller Dateien der teilnehmenden Clients
- Query/Reply
 - Client fragt (Query) nach Dateinamen
 - Server sucht nach passenden Teilnehmern
 - Server antwortet (Reply), wer die Datei besitzt
 - Anfrage-Client lädt Datei vom dateibesitzenden Client herunter



Wie gut ist Napster?

- Vorteile

- Napster ist einfach
- Dateien werden schnell und effizient gefunden

- Nachteile

- Zentrale Struktur erleichtert Zensur, feindliche Eingriffe und technische Pannen
 - wie z.B. Denial-of-Service-Angriff
- Napster skaliert nicht
 - d.h. mit zunehmender Teilnehmerzahl verschlechtert sich die Performanz
 - Speicher auf dem Server endlich

- Resümee

- Napster keine akzeptable Peer-to-Peer-Netzwerklösung
- Bis auf den Download-Aspekt ist Napster im eigentlichen Sinne kein P2P-Netzwerk

Gnutella - Geschichte

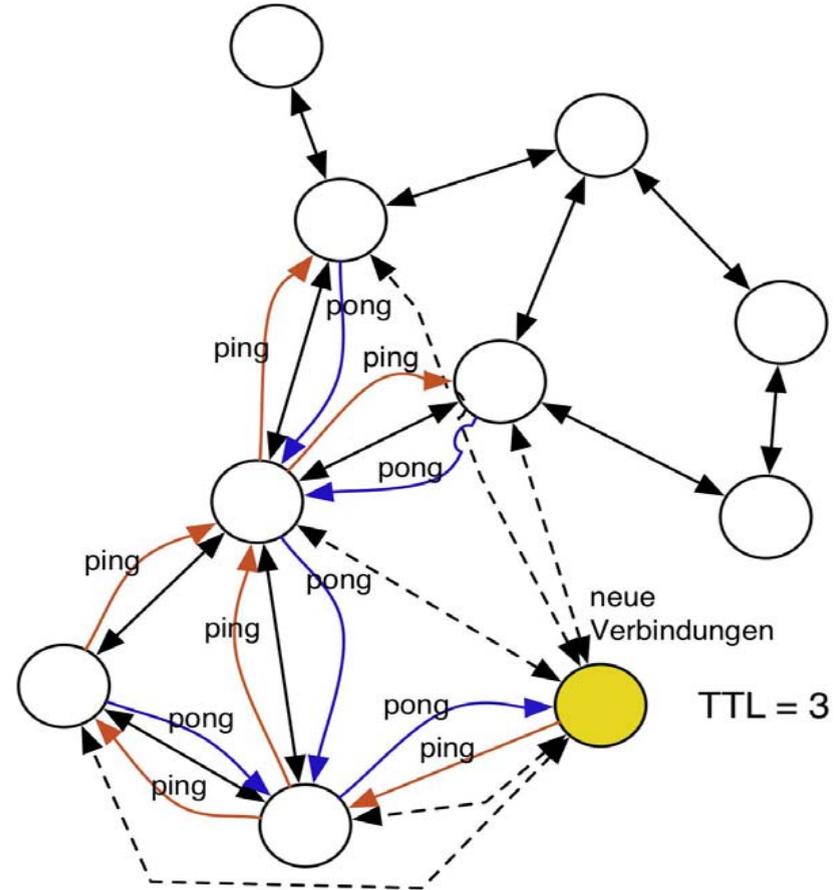
- Gnutella
 - wurde im März 2000 herausgegeben von Justin Frankel und Tom Pepper von Nullsoft
 - Nullsoft ist seit 1999 eine Tochter von AOL
- File-Sharing-System
 - Ziel wie Napster
 - Arbeitet aber völlig ohne zentrale Strukturen



Gnutella - Originalversion - Anbindung

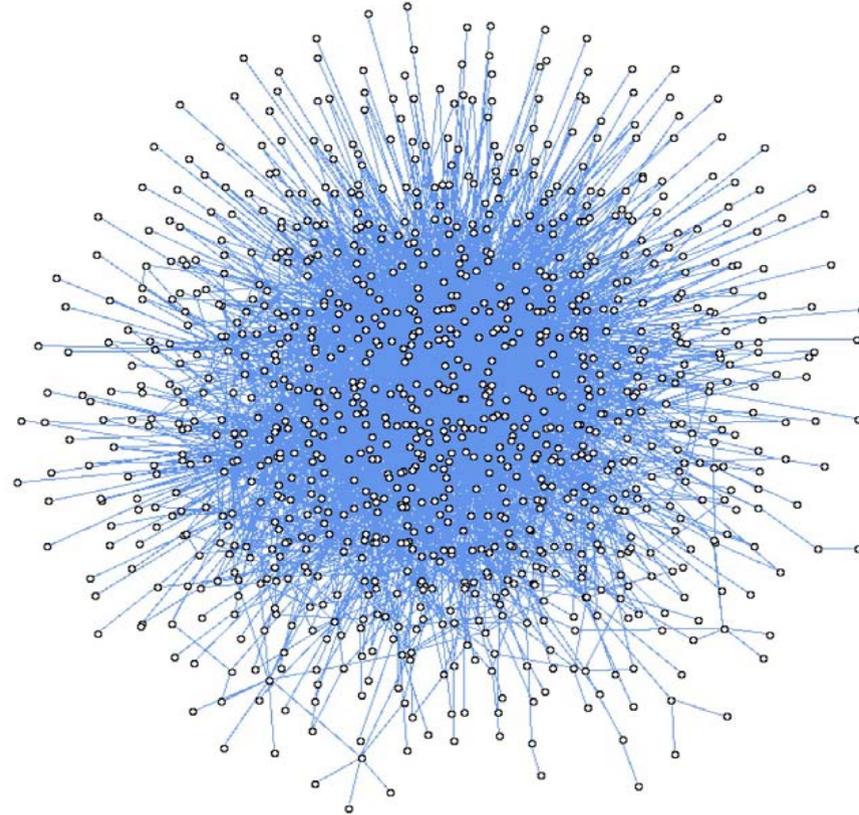
- Nachbarschaftslisten

- Gnutella verbindet direkt mit anderen Clients
- Beim Download wird eine Liste von Clients mitgeliefert
- Diese werden ausprobiert, bis ein aktiver Client sich meldet
- Ein aktiver Client gibt dann seine Nachbarschaftsliste weiter
- Nachbarschaftslisten werden immer weiter verlängert und gespeichert
- Die Anzahl aktiver Nachbarn ist beschränkt (typisch auf fünf)



Gnutella - Originalversion - Anbindung

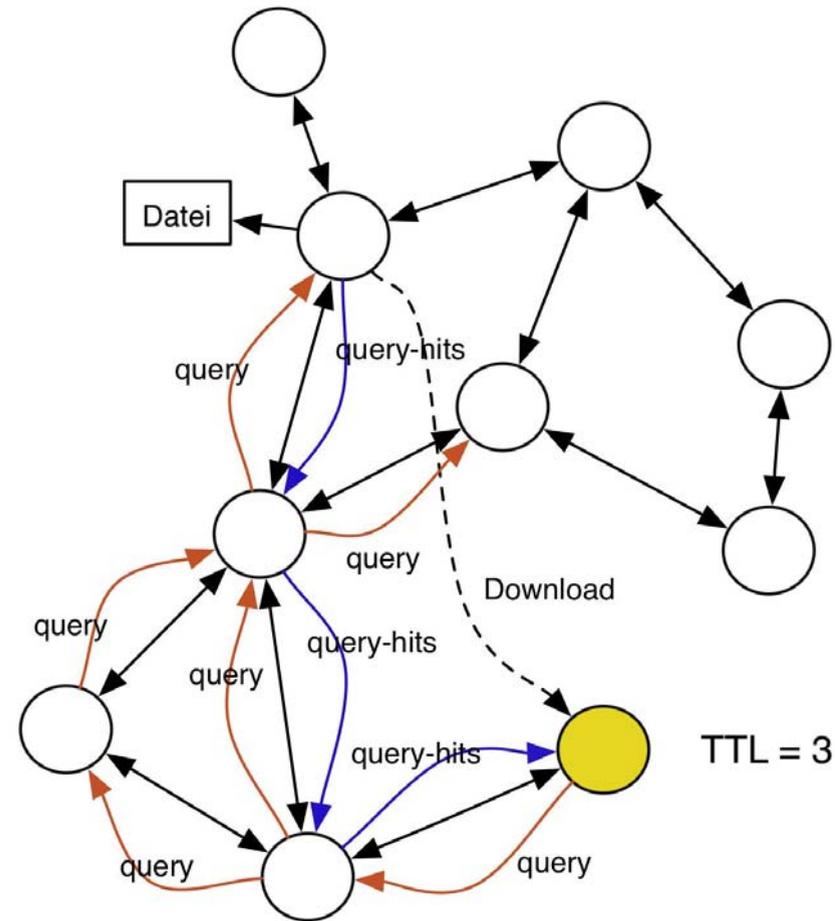
- Protokoll
 - **Ping**
 - Teilnehmeranfrage
 - werden weiter gereicht gemäß TTL-Feld (time to live)
 - **Pong**
 - Reaktion auf Ping
 - Werden auf dem Anfragepfad zurückgereicht
 - IP und Port des angefragten Teilnehmers
 - Anzahl und Größe zur Verfügung gestellter Dateien
- Graphstruktur
 - entsteht durch zufälligen Prozess
 - Knotengrade unterliegen Pareto-Verteilung
 - entsteht unkontrolliert



Gnutella Schnappschuss im Jahr 2000

Gnutella - Originalversion - Anfrage

- Dateianfrage
 - wird an alle Nachbarn geschickt
 - diese senden sie an ihre Nachbarn
 - bis zu einer vorgegebenen Anzahl von Hops
 - TTL-Feld (time to live)
- Protokoll
 - **Query**
 - Anfrage nach Datei wird bis zu Anzahl TTL-hops weitergereicht
 - **Query-hits**
 - Antwort auf umgekehrten Pfad
- Wenn Datei gefunden wurde, direkter Download

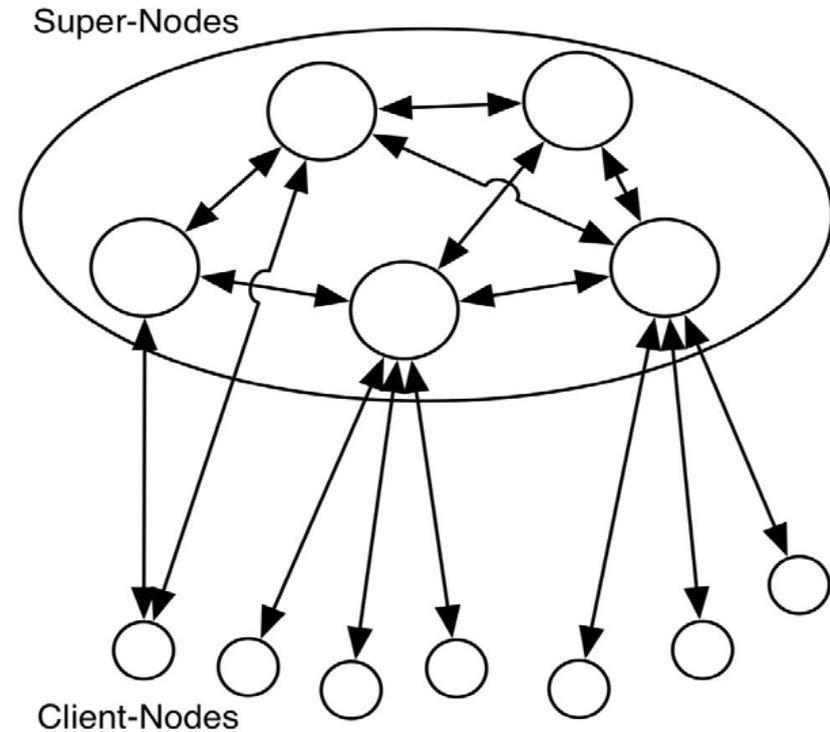


Gnutella - Diskussion

- Vorteile
 - verteilte Netzwerkstruktur
 - Netzwerk (aber nur das) skalierbar
- Nachteile
 - Durch TTL findet für Abfragen eine implizite Netzwerkpartitionierung statt
 - Dadurch Anfrageerfolg gering (!)
 - Durch lange Wege, große Latenzzeiten
- Verbesserungsvorschläge
 - Random Walks statt Broadcasting (verhindert Flooding)
 - Passive Replikation von Information entlang des Pfads
 - Häufigkeit der Replikate nimmt im Quadrat des Abstands ab

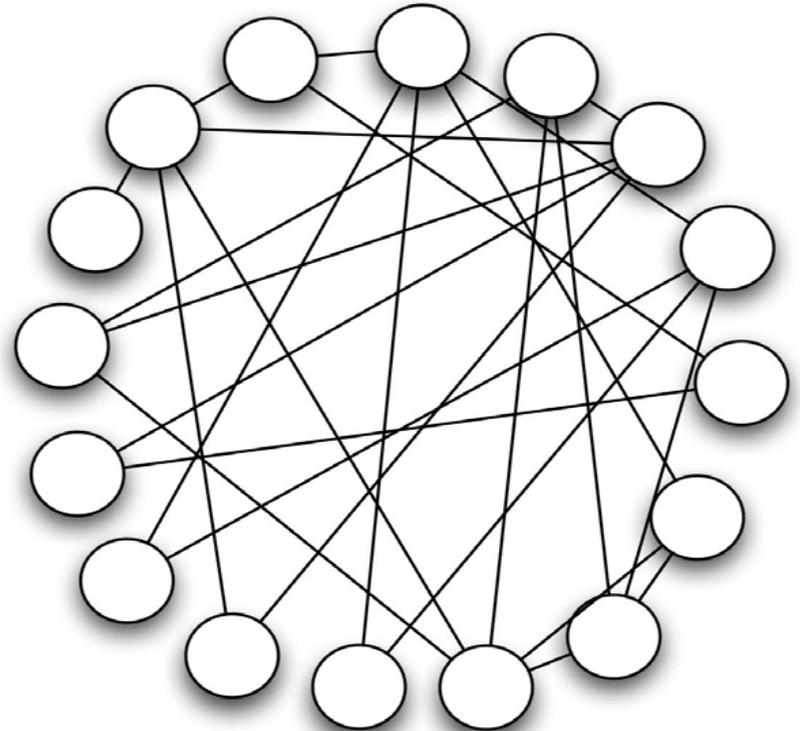
FastTrack & Gnutella2

- Hybride Struktur
 - Knoten mit großer Bandbreite werden zu P2P-Server ausgewählt
 - Diese unterhalten P2P-Netzwerk im Stil von Gnutella
 - Normale Knoten werden als Clients an diese Super-Nodes angebunden
- Eingesetzt in
 - FastTrack
 - Gnutella2 (neuere Ausgabe)
- Vorteile
 - Verbesserte Skalierbarkeit
 - Geringere Latenzzeiten
- Nachteile
 - Immer noch unzuverlässig und langsam
 - Clients können sich der Super-Node-Aufgabe verweigern



Warum skaliert Gnutella nicht?

- Gnutella
 - Graph-Struktur ist zufälliger Verbindungsgraph
 - Grad des Graphen klein
 - Durchmesser gering
 - Zusammenhang groß
- Suche aber aufwendig
 - Um ein Datum sicher zu finden, müßte das gesamte Netzwerk durchsucht werden
- Gnutella skaliert nicht, weil
 - Keine Struktur in der Datenablage

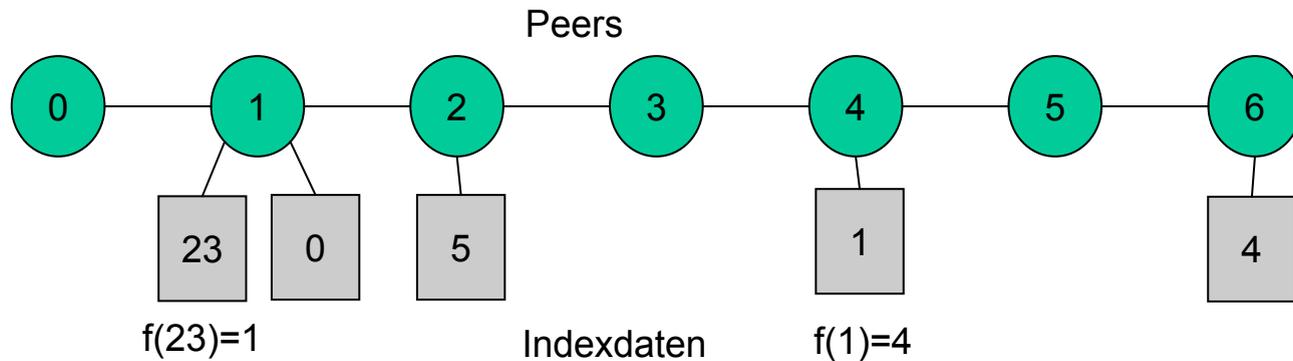


Zwei Fragen zur Informationsfindung

- Wo ist es?
- Wie dorthin kommen?
- Napster:
 - Wo?
 - Auf dem Server ☺
 - Wie dorthin?
 - Zum Serverstau ☹
- Gnutella
 - Wo?
 - Weiss nicht ☹
 - Wie dorthin?
 - Alle fragen ☹
- Besser:
- Wo ist Datum x ?
 - An der Stelle $f(x)$
 - Was ist $f(x)$?
 - Eine allen Teilnehmern bekannte Abbildung von x auf einem Raum
- Wie komme ich dorthin?
 - Durch eine Route die mir den Weg von meinem Standort zu $f(x)$ aufzeigt.

Eine Hash-Tabelle als Peer-to-Peer-Netzwerk

- Jeder Peer steht für eine Speicherstelle $0, 1, 2, \dots, n-1$
 - Eine allen Peers bekannte Hash-Funktion, z.B. für $n = 7$ Peers
 - $f(x) = (3x+1 \bmod 23) \bmod 7$
 - Peers sind als Kette verbunden

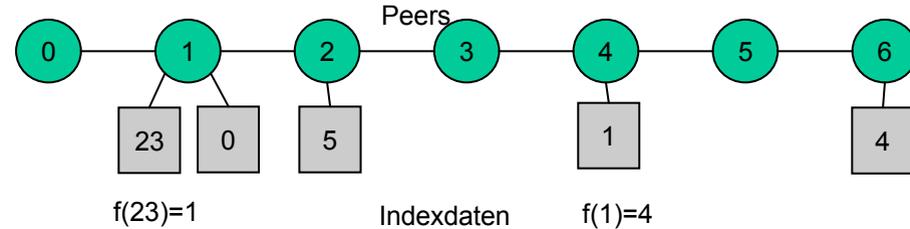


- Suche
 - Berechne $f(x)$
 - Gehe zu Peer mit Adresse $f(x)$ entlang der Linie

Distributed Hash-Table (DHT)

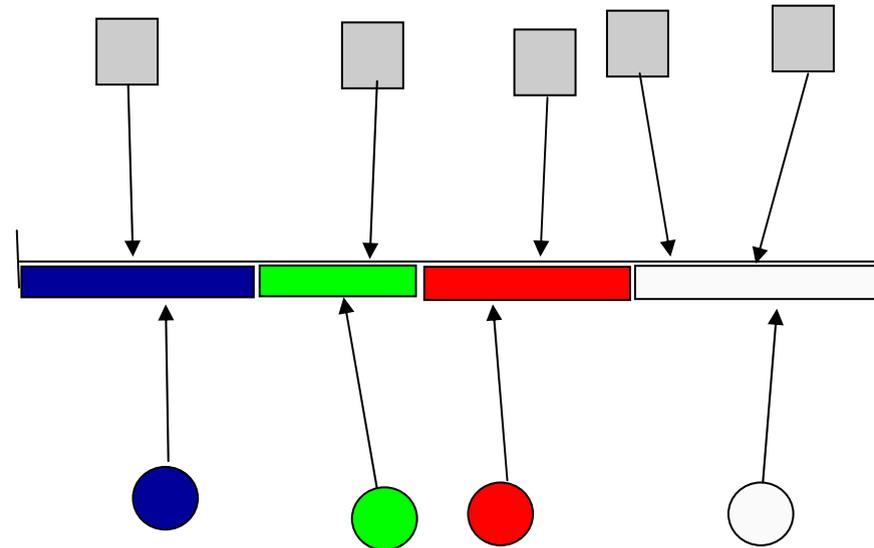
Hash-Tabellen

- Vorteile
 - Suche einfach
- Nachteile
 - Ein neuer Peer verursacht die Wahl einer neuen Hash-Funktion
 - Lange Wege



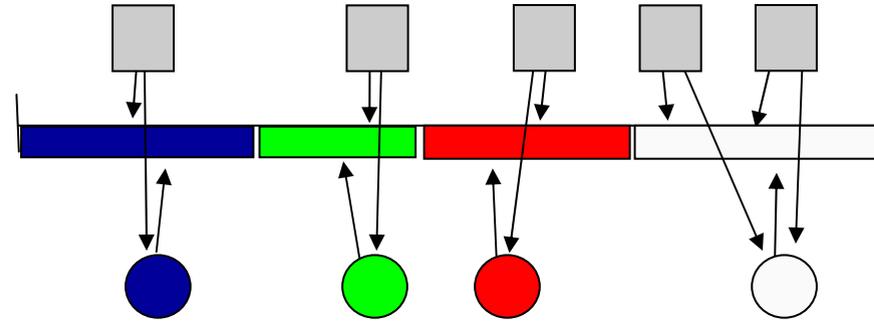
Distributed Hash-Table

- Peers werden an eine Stelle ge„hash“t und erhalten Bereiche des Wertebereichs der Hashfunktion zugeteilt
- Daten werden auch ge„hash“t
 - Je nach Bereich den Peers zugeordnet

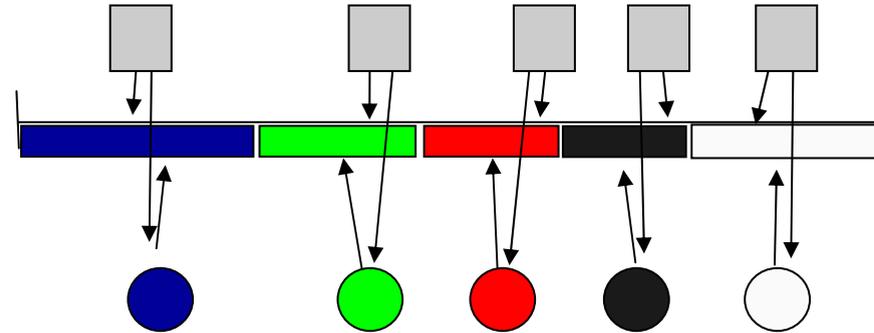


Einfügen in die Distributed Hash-Table (DHT)

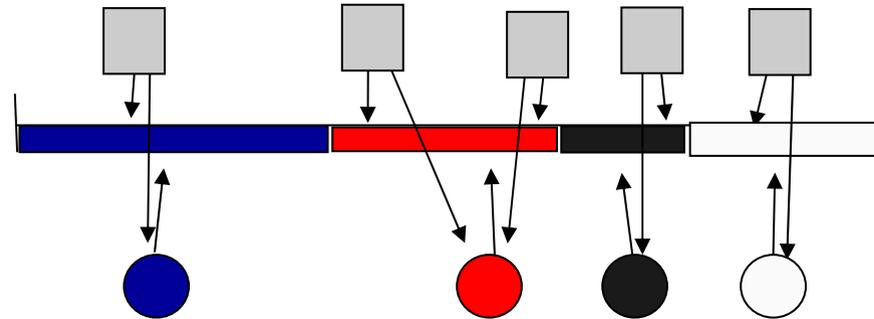
- Distributed Hash-Table
 - Peers werden an eine Stelle ge„hash“t
 - Dokumente ebenso
 - Jeder ist für einen Bereich verantwortlich



- Kommt ein neuer Knoten hinzu
 - müssen die Nachbarn teilen

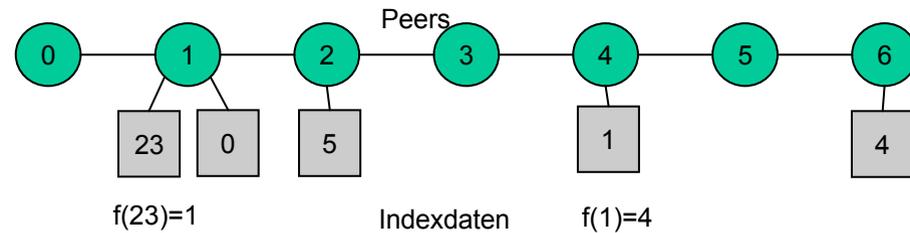


- Verlässt ein Knoten das Netzwerk
 - übernehmen die Nachbarn sein Gebiet



Eigenschaften DHT

- Vorteile
 - Jedes Datum kann einem bestimmten Peer zugewiesen werden
 - Einfügen und Entfernen von Peers erzeugt nur Veränderungen in den benachbarten Peers



- DHTs werden von vielen P2P-Netzwerken benutzt
- Noch zu klären:
 - Die Verbindungsstruktur

