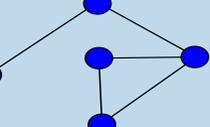
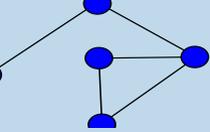

Vorlesung: P2P und Datenbanken

Dr. Dieter Sosna

7. April 2008

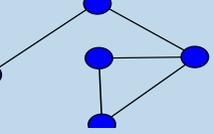


Vorbemerkung



- Vorlesung - 2 SWS
 - im Rahmen des Moduls „Anwendungsbezogene Datenbankkonzepte“ (Master neuer Ordnung),
 - Master (alte Ordnung),
 - Diplomstudiengang „Informatik“,
 - Hörer anderer Studiengänge
- Erwartete Vorkenntnisse: Vorlesung DBS-1 (u. -2)
- Abschlußprüfung (Klausur) im Juli 2008.

Jede **Teilnahmebestätigung** setzt grundsätzlich **erfolgreiche Teilnahme an der Prüfung** voraus.
- Prüfungsrelevant ist die Vorlesung, nicht das Skript !

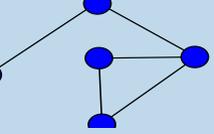


Kapitel 1: Einführung

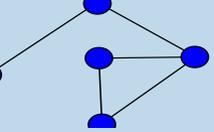
Vorbemerkung

Definition u. Merkmale

Einsatz - allgemein

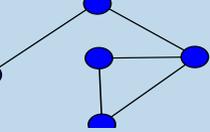


Definition u. Merkmale



Kapitel 1: Motivation

- Einschätzungen
- Einsatzgebiete
- Charakteristische Merkmale und Typen
- Das Internet - ein P2P-Netz ?



Clay Shirky :

<http://www.openp2p.com/pub/a/p2p/2001/04/05/shirky.html>

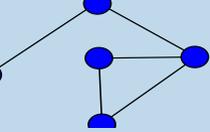
... The Wall St. Journal's take on peer-to-peer is simple and direct: it's not making investors any money right now. ...

...Gomes: – P2P isn't a well-defined business sector, nor a well-defined technology – are both sensible. From a venture capitalist's point of view, P2P is too broad a category to be a real investment sector.

The unspoken premise of both articles is this:

if peer-to-peer is neither a technology or a business model, then it must just be hot air. There is, however a third possibility besides „technology“ and „business“. The third way is simply this:

Peer-to-peer is an idea.



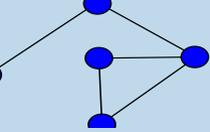
Definition

(nach Clay Shirky)

Definition:

Mit Peer-to-Peer bezeichnen wir eine Klasse von **Applikationen** (auch im Sinne des ISO-Schichtenmodells) , die

- Ressourcen benutzen, welche im Internet verfügbar sind: Speicher, Rechenleistung, Inhalte, Leistung von Menschen.
- dabei ein logisches Netz bilden, für das charakterisch ist,
 - daß die Verbindungen zeitlich nicht stabil sind,
 - Verfügbarkeit einzelner Teilnehmer nicht garantiert wird,
 - IP-Adressen nicht vorhersagbar sind (DNS nicht nutzbar).
 - es keine zentrale Autorität gibt



Test-Kriterium

(nach Clay Shirky)

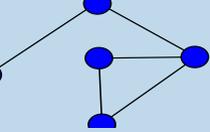
Abgrenzung zu verteilter Anwendung? P2P „litmus“-Test:

- Erlaubt die Anwendung temporale Variabilität bei der Teilnahme an dem logischen Netzwerk, werden zeitlich variable Adressen des darunterliegenden Internets unterstützt?
- Erlaubt die Anwendung ein hinreichend großes Maß an Autonomie der Knoten (ohne daß die Funktionalität des Gesamtsystems Einbusen erleidet)?

Ermessensspielraum bei der Bewertung:

Wie kann Autonomie gemessen werden?

Abgrenzung zu Mißbrauch? Lassen sich für Teilbereiche quantitative Maße finden?

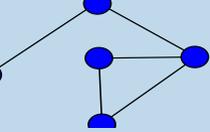


Charakteristische Merkmale

- Die Peer-to-Peer-Anwendung hat ihrer Natur nach eine *Client-Server-Struktur -aber eine spezielle.*
- Jeder Knoten des logischen Netzes ist **sowohl Server als auch Client.**
- Alle Peers sind autonom (Sie entscheiden autonom über Teilnahmezeit, -ort, z.T. über angebotene Services)
- Peers und Verbindungen sind unsicher
- Alle Peers sind gleichberechtigt (Namensgeber für diese Klasse von Anwendungen)
Jeder Peer kann alle existierenden Daten und Services nutzen,
jeden anderen Peer kontaktieren,
muß/soll auch alle Dienste anbieten.

Logisches Netz in Anwendungsschicht

(Die darunter liegenden Schichten bleiben transparent - Beispiel: echolink (s.später)



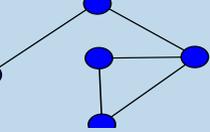
Charakteristische Merkmale (2)

- Es existiert keine zentrale Koordinierung u. keine zentrale Datenbank (1)
- Kein Peer hat eine globale Übersicht über das System (2)
- Das globale Verhalten bestimmt sich aus der Gesamtheit aller lokalen Beziehungen zwischen zwei Peers.

Einige Anwendungen erfordern Systemeigenschaften, die mit diesen Merkmalen (1)(2) nur schwer oder derzeit nicht erreichen lassen.

Definition:

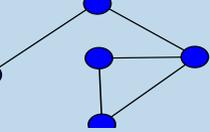
- Sind alle Knoten in ihrer Funktionalität untereinander gleich, liegt eine **reine P2P-Anwendung** vor.
- Sind bei einem oder einigen Knoten die Serverfunktionen stärker oder anders ausgeprägt als bei anderen, ergibt sich eine **hybride P2P-Anwendung**.



Charakteristische Merkmale (3)

Hybride Systeme

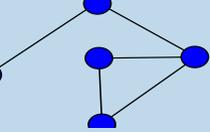
- Bei einigen Anwendungen sind die geforderten Eigenschaften leichter bzw. bei derzeitigen Stand nur in hybriden Systemen erreichbar.
- Zwei Typen hybrider Systeme:
 - P2P-Netze mit zentralem Server:** Wichtige Funktionen (z.B. globaler Index, Mitglieder, ...) auf dem Server zentralisiert. \mapsto sensibler Punkt. Server vermittelt direkten Kontakt der Peers.
Beispiel: Napster, Echolink
 - Hierarchische Netze:** (i.A. mehrere) Superpeers mit herausgehobener Funktion, Mischarchitektur.



Klassifikation

- strukturiert: Platzierung der Daten nach festgelegten Plan.
 - Plan kann bei Lokalisierung ausgenutzt werden
- unstrukturiert: Zufällige Verteilung.
 - z.B. beim Autor
- semistrukturiert: Plan mit Zufall kombiniert,
 - z.B. Regeln zur Datenmigration in Richtung Nachfrager
 - Geheimnisteilung

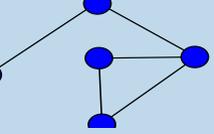
Datenverteilung hat Folgen für das System: Anonymität, Sicherheit gegen Angriffe auf Verfügbarkeit, Update-Fähigkeit, ..



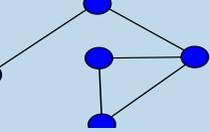
Klassifikation

- strukturiert:
 - ein oder mehrere (komplexe) Schemata,
 - Inhalte feingranular, getypt.Die Struktur wird vom System erkannt und ausgenutzt.
- unstrukturiert:
 - die kleinste sinnvoll verarbeitbare Einheit ist die Resource oder
 - evt. vorhandene Strukturen werden nicht benutzt
- semistrukturiert:

Merkmale von strukturierten und unstrukturierten Ressourcen vorhanden
Besserer Begriff: partiell strukturiert.



Einsatz - allgemein

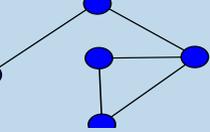


Einsatzgebiete

Literatur: http://www.openp2p.com/pub/q/p2p_category

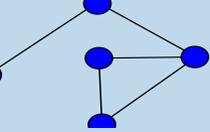
P2P-Systeme in verschiedenen Gebieten möglich:

- Verteilte Objekte (Daten, Dateien - Filesharing bisher dominierend, Probleme: Lokalisation, Identifikation der Daten, Versionsverwaltung, free-riding, Netzlast)
- Verteiltes Rechnen (SETI@home , nicht zeitkritisch, kooperativer Peer vorausgesetzt)
- andere WEB-Services (e-commerce)
- Devices als Peers (Bluetooth)
- Zusammenarbeit, Wissensverwaltung (Authentifikation)
- Verteilte Entwicklungsarbeit
- Mobile Ad-hoc Netzwerke
- ...



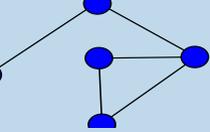
Erweiterungspotential

- Strukturierte Daten:
 - Techniken zur Daten- und Schemaintegration
 - Übertragung von Datenbanktechnologien
 - Kann es ein Transaktionskonzept geben?
 - Geschäftsprozesse
 - Identifikation und Authentifizierung der Partner bei evt. notwendiger Geheimhaltung der Beziehung.
 - Qualität der angebotenen Dienste (inhaltlich, temporal)
 - Qualitätssicherung bei Mehrschrittaktionen (Dienstleistung und Bezahlung)
- Bieten *mobile Agenten* eine Lösung einiger Probleme?



Historisches

- Das originale Konzept des Internet hatte P2P-Konzepte
Literatur: Michael Hauben: History of ARPANET, URL: www.dei.isep.ipp.pt/docs/arpa.html
Reaktion auf Sputnikstart, erste Realisierung 1968
Direkter Kontakt zweier Rechner, um Ressourcen zu teilen
WAN zuerst zwischen 4 militär. Einrichtungen, später auch zwischen Universitäten, Forschungseinrichtungen.
Schaffung der Standards war ein hochkomplexes Entwicklungsprojekt.
- Wenn ein Teil des Netzes ausfällt, kann der Rest weiterarbeiten.
(Beispiel: Konfiguration des Netzzugangs eines Rechners - mehrere Nameserver möglich, ...)
- Auch C/S-Anwendungen (ftp), z.B. kann jeder Server betreiben.



Zusammenfassung

- Altes Arbeitsparadigma auf Anwendungsniveau neu realisiert.
- Symmetrische Client-Server-Architektur
- Hochgradige Autonomie der Teilnehmer, hochgradige Fluktuation, instabile Verbindungen
- Untertypen: rein, hybrid (mit zentralem Server, hierarchisch)
- Klassifikation der Datenverteilung, der Ressourcen.
- Internet als P2P-System.