

Nutzen von p2p Techniken für die E/A in verteilten Rechnerumgebungen

Autor: Arne Babenhauserheide, Betreuung: Julian Kunkel

28. Oktober 2008

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Gliederung

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

In diesem Vortrag bespreche ich die Eignung der Dateivervielfältigung via BitTorrent für Cluster- und WLAN-Umgebungen.¹²

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

¹Ich stelle die Ergebnisse des Praktikums vor, nicht seinen Ablauf im Einzelnen, da mir die Ergebnisse interessanter erscheinen. Der Ablauf ist in meinem Bericht zu finden.

²Ich spreche hier als 'ich', weil ich das Praktikum in Einzelarbeit gemacht habe. Allerdings hat Julian mir verdammt gute Ratschläge und Tipps gegeben, ohne die ich das Praktikum vermutlich nicht in der Qualität hatte abschließen können, die ich von mir selbst erwarte. 

Warum p2p Techniken - aktuelle Situation im Hochleistungsrechnen

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

- ▶ Flaschenhals: Nicht länger Netzwerk, sondern Einzel-E/A.
- ▶ Dateivervielfältigung in Clustern entspricht der Situation von p2p Netzen.
- ▶ Außerdem: Dateivervielfältigung für WLAN meshes.

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Verteilte gegenüber zentralisierter Dateivervielfältigung

- ▶ Zentralisiertes System: Einige wenige Dateiserver versorgen alle.
- ▶ Dezentral: Der Großteil der Last wird auf die Knoten verteilt.
- ▶ Skalierung: Datenserver: $O(1)$, p2p Techniken: $O(N)$.

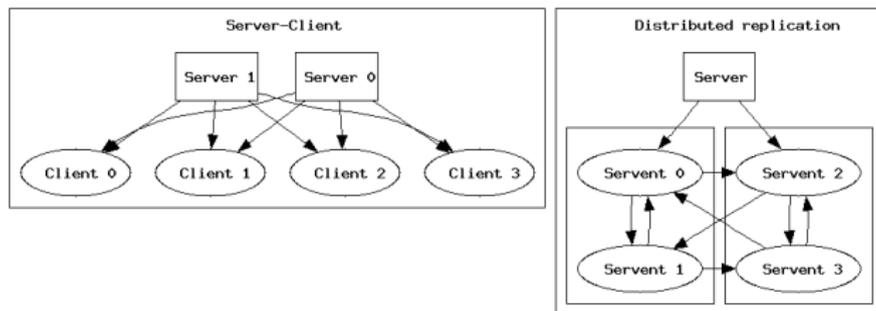


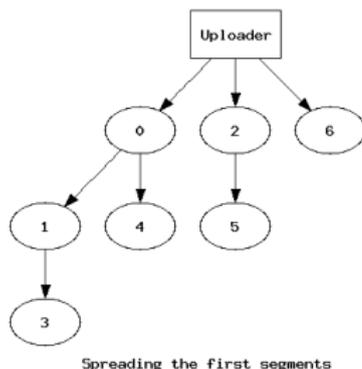
Abbildung: Gegenüberstellung: Ein verteiltes Dateisystem mit mehreren Datenservern und ein System mit verteilter Dateivervielfältigung

Schematik für verteilte Vervielfältigung

- ▶ “Knoten”: Eine einzelne Einheit mit eigenem Speichersystem.
- ▶ “Erste Quelle”: Der Knoten, von dem aus die Datei vervielfältigt wird.
- ▶ “Startphase”: Die Zeit, bis jeder Knoten mindestens ein Fragment der zu replizierenden Datei hat.
- ▶ “Endphase”: Die Zeit, in der die erste Quelle das letzte Fragment weitergegeben hat, aber noch nicht alle Knoten alle Fragmente haben.
- ▶ “Hauptphase”: Die Zeit, in der alle Knoten ihre volle Netzwerkleistung zur Verfügung stellen.
- ▶ : “N”: Anzahl der Knoten

Grundlegende Leistungsbetrachtung

- ▶ Startphase und Endphase: Dauer mindestens $\log(N)$ -Zeit für ein Fragment, wenn die erste Quele die Daten nur genau einmal hochlädt.
- ▶ Maximalgeschwindigkeit in Hauptphase: Anzahl der Knoten \cdot Einzelgeschwindigkeit, aber höchstens Gesamtgeschwindigkeit des Netzes.



Structure of the first fragments spreading in the network when a transfer gets initiated.

Abbildung: Startphase: Die ersten Fragmente verbreiten

Maximale Übertragungsrate in Startphase

$$\sum_{i=1}^{\log_2(N)} (2^{i-1} / \log_2(N)) \cdot \text{Einzelrate}(\text{upload})$$

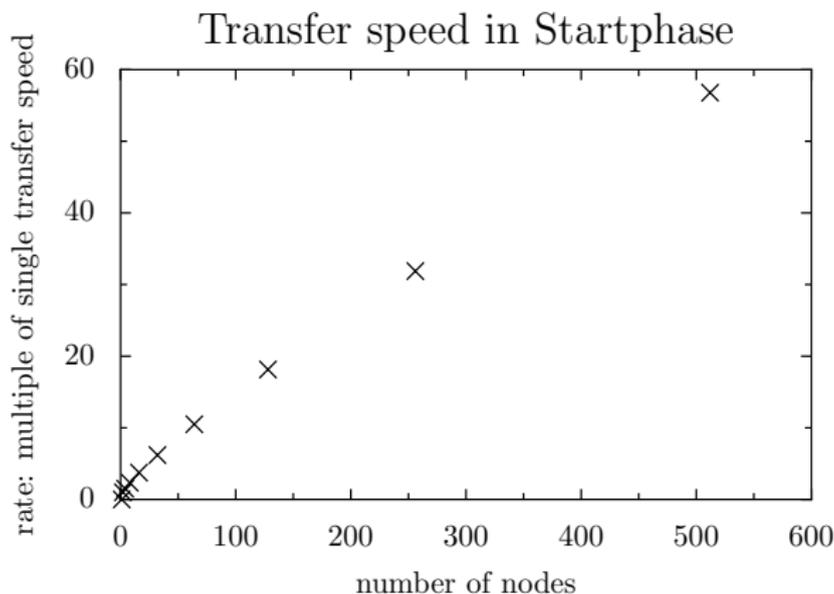


Abbildung: Effektive Übertragungsrate in der Startphase bei Netzen mit unterschiedlich vielen Knoten

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

[Einstieg](#)

[Warum und Was?](#)

[Leistungsmessung](#)

[Abschluss](#)

Maximale Übertragungsrate der gesamten Übertragung

Maximale Übertragungsrate:

$$s_{ges} = \frac{\text{Datenmenge}}{2 \cdot \frac{\log_2(7) \cdot \text{Fragmentgrösse}}{\text{Plattenleistung}} + \frac{(\text{Datenmenge} - 2 \cdot \text{Anfangsrate} \cdot \text{Anfangszeit})}{\min(7 \cdot \text{Plattenleistung}, 380 \text{ MiB/s})}}$$

Für unseren Cluster:

$$s_{ges} = \frac{816 \text{ MiB}}{2 \cdot 0.036 \text{ s} + \frac{816 \text{ MiB} - 2 \cdot 0.083 \text{ MiB}}{294 \text{ MiB/s}}}$$

$$s_{ges} = 286.26 \text{ MiB/s}$$

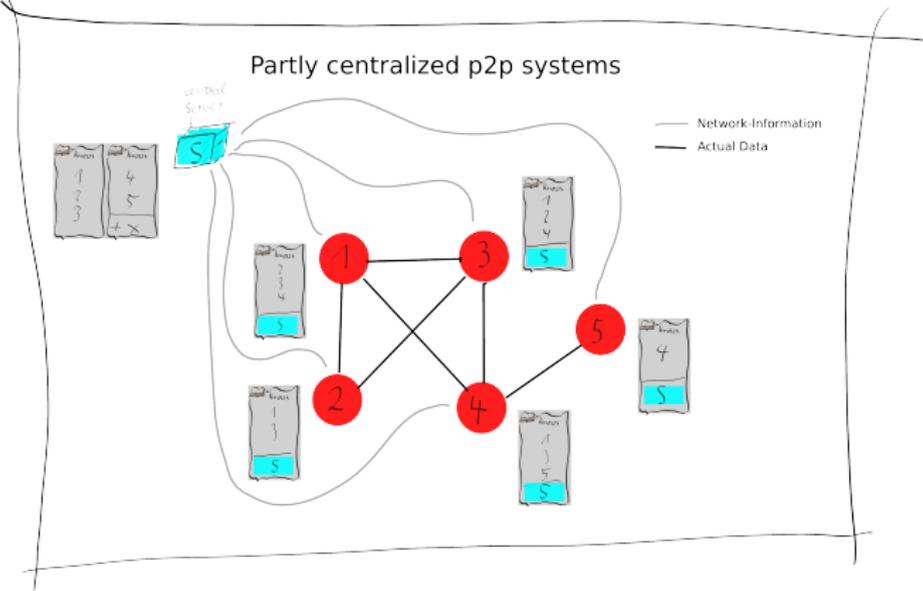
Daten:

- ▶ Plattenleistung: 42MiB/s
- ▶ Max Netz: 380MiB/s
- ▶ Anzahl Knoten: 7
- ▶ Fragmentgröße: 512kiB/s

Das untersuchte System: BitTorrent

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel



- Einstieg
- Warum und Was?
- Leistungsmessung
- Abschluss

Abbildung: Struktur von teilzentralisierten p2p Netzwerken

Das untersuchte System: BitTorrent

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

- ▶ BitTorrent ist teilzentralisiert,
- ▶ benötigt für jede Übertragung einen Koordinierungsserver,
- ▶ ist leicht aufzusetzen und
- ▶ industriell großteils anerkannt.

Komplett dezentrale Systeme bespreche ich zum Abschluss für ausgedehnte WLAN-Umgebungen noch einmal.

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Ob BitTorrent im Cluster hält, was es im Internet verspricht habe ich in unserem Testcluster untersucht. Um seine Leistung testen zu können, habe ich eine kleine Testumgebung geschrieben, mit der ich verschiedene automatische Tests laufen lassen konnte.

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Die Testumgebung

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

- ▶ BitTornado als Server, rtorrent für die Knoten.
- ▶ Python Skripte zur Kontrolle.
- ▶ Sehr einfache Möglichkeit, Szenarien zu definieren.
- ▶ Geschwindigkeitskontrolle durch rtorrent Konfigurationsdateien.

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Szenariodefinition

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

```
'''Call this scenario via:
```

```
>>>scenario(run)'''
```

```
-
```

```
def scenario(run, logfile='logfile.log'):
```

```
    PATH = '/tmp/Infinite-Hands-free-software.ogg'
```

```
    NUMBER_OF_TESTS = 30
```

```
    WAIT_TIME = 120
```

```
    CONFIG = 'rtorrent-WLAN.rc'
```

```
-
```

```
    run(path=PATH, number_of_tests=NUMBER_OF_TESTS,  
        wait_time=WAIT_TIME, config=CONFIG, logfile=logfile) 3
```

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

³Ohne den Großteil der Kommentare für [Lesbarkeit auf Folien.](#)

rtorrent Konfigurationsdatei

Extrem einfach:

```
.rtorrentrc:
```

```
upload_rate = 2875
```

```
download_rate = 28754
```

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

⁴In kiB - viel einfacher geht eine Geschwindigkeitsbegrenzung kaum 

Gewählte Szenarien

Geschwindigkeiten:

- ▶ Ohne Begrenzung: max. 380MiB/s - Netz des Clusters
- ▶ WLAN Umgebung: 23 Mbit - 2875kiB/s⁵



Abbildung: Zwei Netzwerk-Arten: Kabel + WLAN

⁵OLPC Meshes haben 56Mbit, also das doppelte des getesteten Wertes.

Gewählte Szenarien

Vervielfältigte Daten:



Abbildung: Vervielfältigte Daten

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Testergebnisse - WLAN

Das interessanteste zuerst: WLAN

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

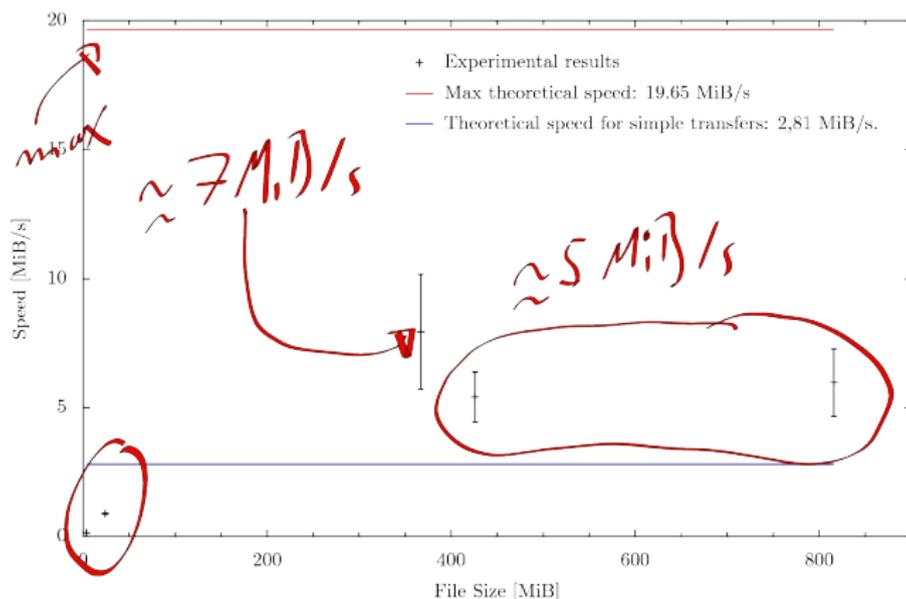


Abbildung: Geschwindigkeit der Übertragung bei WLAN
Bandbreite

Testergebnisse - WLAN

Offset Bestimmung:

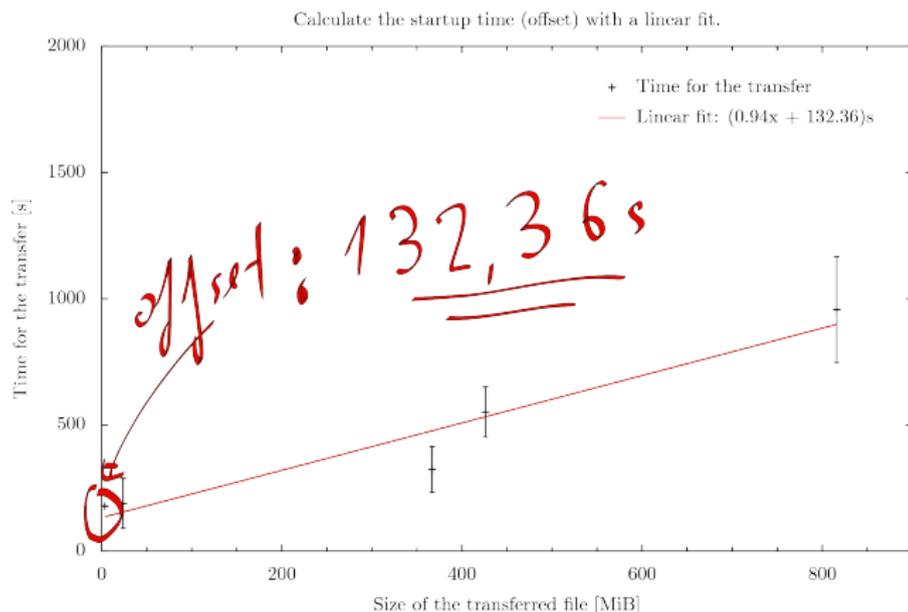


Abbildung: Offset-Bestimmung Übertragungszeiten - WLAN

Testergebnisse - WLAN

Offset-korrigierte Geschwindigkeit:

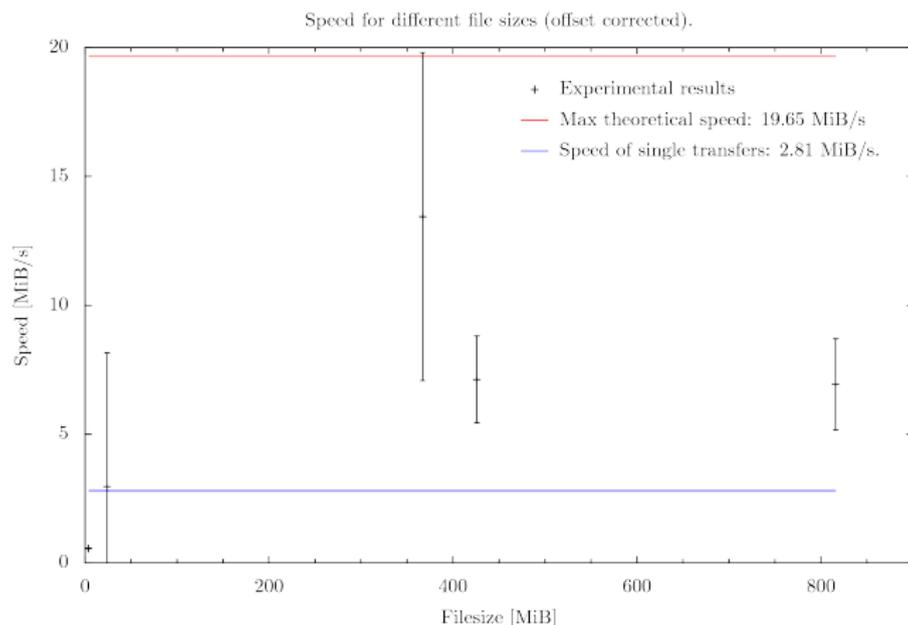


Abbildung: Offset-korrigierte Geschwindigkeit der Übertragung bei WLAN Bandbreite

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

[Einstieg](#)

[Warum und Was?](#)

[Leistungsmessung](#)

[Abschluss](#)

Testergebnisse - WLAN

Offset-korrigierter Speedup:

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

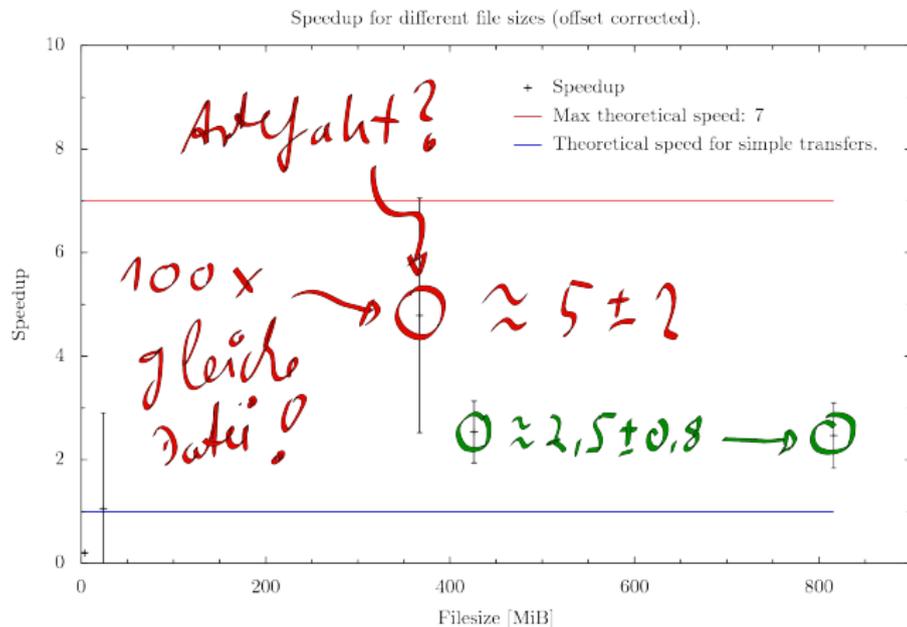
Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Einstieg

Warum und Was?

Leistungsmessung

Abschluss



captionOffset-korrigierter Speedup der Übertragung bei
WLAN Bandbreite

Testergebnisse - WLAN

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Als Ergebnis für WLAN habe ich die Übertragungszeit mit BitTorrent nocheinmal linear gefittet und dazu die effektive Übertragungsgeschwindigkeit bei Einzelübertragungen aufgetragen.

An dem Vergleich wird deutlich, dass BitTorrent bei WLAN Geschwindigkeiten mit 7 Downloadern ab etwa 100MiB zu übertragenden Daten effizienter wird als Einzelübertragungen (hier ist der Offset noch nicht korrigiert, da die Offset Korrektur nur zur Abschätzung der Effizienz von BitTorrent in größeren Clustern dient).

[Einstieg](#)

[Warum und Was?](#)

[Leistungsmessung](#)

[Abschluss](#)

Testergebnisse - WLAN

Vergleich des Zeitbedarfs, bis alle Übertragungen abgeschlossen sind, Einzelübertragungen gegenüber BitTorrent-Übertragung.

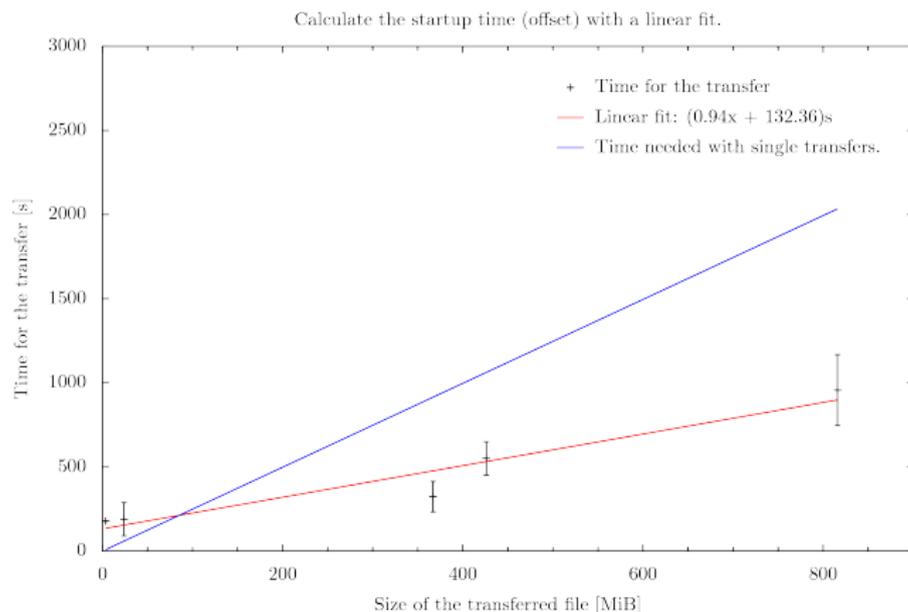


Abbildung: Vergleich der benötigten Zeit mit Einzelübertragungen und mit BitTorrent ohne Offset-Korrektur

Testergebnisse - volle Geschwindigkeit

Bei voller Geschwindigkeit habe ich nur mit dem 816MiB Film getestet. Selbst mit Offset erreicht BitTorrent gerade mal 1/4 der erwarteten Bandbreite. D.h. es sind noch einige Optimierungen nötig.

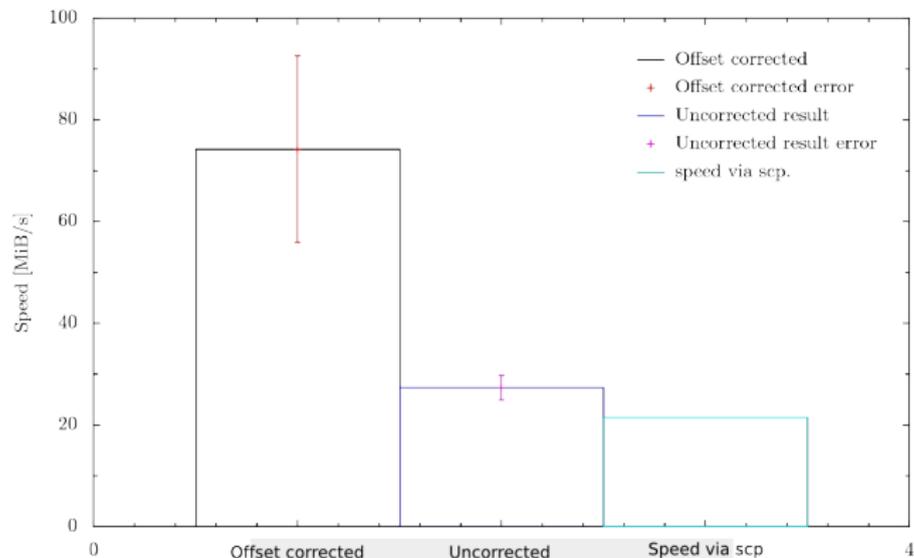


Abbildung: Geschwindigkeit der Übertragung bei unbegrenzter Bandbreite

Testergebnisse - Fazit

- ▶ Cluster nein,
- ▶ WLAN ja. ⁶

Mögliche Gründe:

- ▶ CPU Last, z.B. durch TCP Verbindungen (mit top beobachtet).
Diese CPU-Last ist im Internet vernachlässigbar, in Clustern wegen der deutlich höheren Paketrage aber ein wichtiger Aspekt.
- ▶ BitTorrent ist für die Nutzung im Internet optimiert.

⁶Drosselalgorithmus von rtorrent nicht vertrauenswürdig genug
- Die Daten sind mit zusätzlicher Unsicherheit behaftet. 

“Wie wir also gesehen haben fällt die Leistung von BitTorrent im Praxistest in Clustern leider hinter die Erwartungen zurück und dürfte sich vor allem in WLAN-Meshes eignen. Daher bespreche ich im letzten Teil des Vortrags kurz was die besonderen Bedingungen in WLAN-Meshes für Dateivervielfältigung bedeuten, und wie eine ideale Dateivervielfältigung in Clustern realisiert werden könnte.”

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Dateivervielfältigung in WLAN-Meshes

Komplett dezentrales System (z.B. Gnutella Download-Mesh).

BitTorrent benötigt immer eine Verbindung zu einem zentralen Koordinierungsserver, was in dynamischen WLAN Mesh-Netzwerken problematisch werden kann.

Echt dezentrale Netzwerke bilden eine Alternative, da hier nur

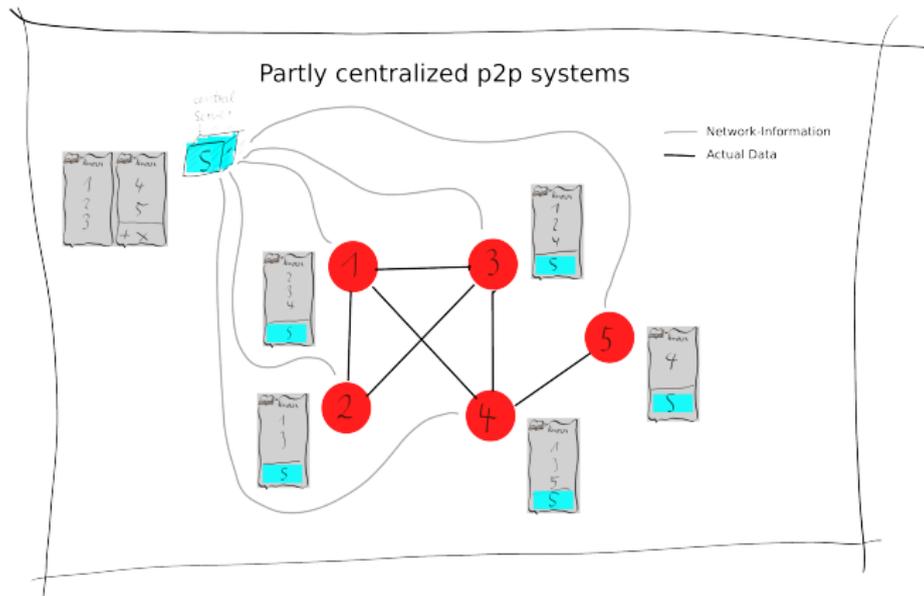
- ▶ bekannt sein muss, dass es eine Datei gibt (+ Hash-ID) und
- ▶ einer der Knoten in Reichweite an der Vervielfältigung teilnehmen muss.

Der Unterschied zwischen BitTorrent und echt dezentralen Netzen wird auf den folgenden zwei Folien visualisiert.

p2p Netzwerktypen - teilzentralisierte

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel



Einstieg

Warum und Was?

Leistungsmessung

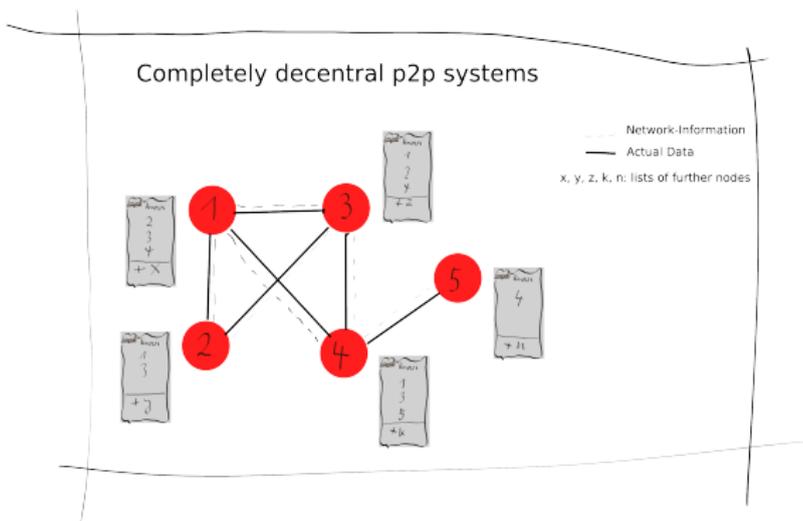
Abschluss

Abbildung: Struktur von teilzentralisierten p2p Netzwerken

p2p Netzwerktypen - echt dezentrale

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel



Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Abbildung: Struktur von vollständig dezentralen p2p Netzwerken

p2p-Netzwerktypen

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Echt dezentrale Netze sind dadurch sinnvoller für Netze,
deren Topologie größtenteils unbekannt ist.

Ihre Leistung könnte z.B. in einem weiteren Projekt getestet
werden.

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Struktur einer möglichen idealen Dateivervielfältigung für Cluster

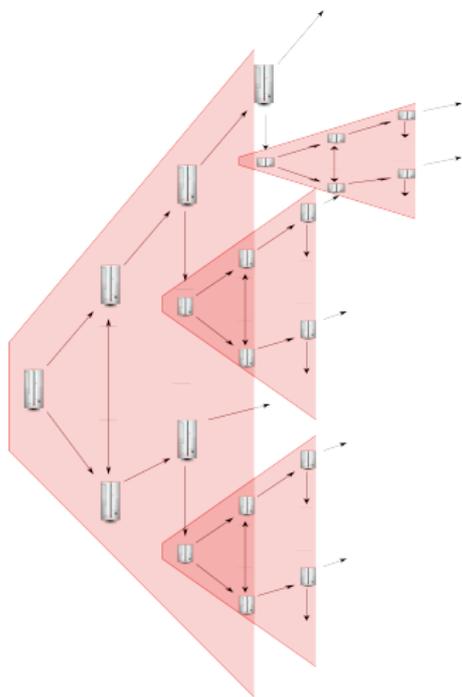


Abbildung: Struktur einer möglichen idealen Dateivervielfältigung für Cluster

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Eigenschaften der Struktur

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

[Einstieg](#)

[Warum und Was?](#)

[Leistungsmessung](#)

[Abschluss](#)

- ▶ Benötigt caching - Nur bei Quelle ist die Festplatte die Grenze, andere Knoten brauchen min. 2x Geschwindigkeit der E/A der Quelle (sonst keine volle Auslastung und nur max. Geschw. = $1/2 \cdot \text{Anzahl Knoten} \cdot \text{Einzelgeschwindigkeit}$)
- ▶ Geschwindigkeit $O(N)^7$.
- ▶ Ausfall eines Knotens legt nur den Teilast lahm, in dem er hängt.
- ▶ Dauer der Start- und Endphase: $O(\log(N))$

⁷Abgesehen von Start- und Endphase

Fast fertig :)

Die gesamte Testumgebung sowie alle generierten Daten und Ergebnisse gibt es auf den Seiten der Gruppe parallele und verteilte Systeme der Universität Heidelberg.

(erstmal hier: <http://draketo.de/dateien/p2p-perf-test.tar.bz2>)

Verwendete Quellen für diesen Vortrag: Icons des Oxygen Projektes (<http://www.oxygen-icons.org/>) unter der GNU Lesser Public License.

Und abschließend: Herzlichen Dank an Julian Kunkel, für viele Inspirationen und für seine Betreuung des Praktikums, und an Prof. Ludwig, dass ich das etwas unkonventionellere Praktikum machen und dafür einen guten Monat lang den Testcluster in Beschlag nehmen durfte!

Abschluss

Nutzen von p2p
Techniken für die
E/A in verteilten
Rechnerumgebun-
gen

Autor: Arne
Babenhauserheide,
Betreuung: Julian
Kunkel

Einstieg

Warum und Was?

Leistungsmessung

Abschluss

Danke für eure Aufmerksamkeit!
Bitte stellt die Fragen, die euch eingefallen sind.