

# A Transactional Scalable Distributed Data Store: Wikipedia on a DHT

**T. Schütt**, M. Moser, S. Plantikow, **F. Schintke**, A. Reinefeld

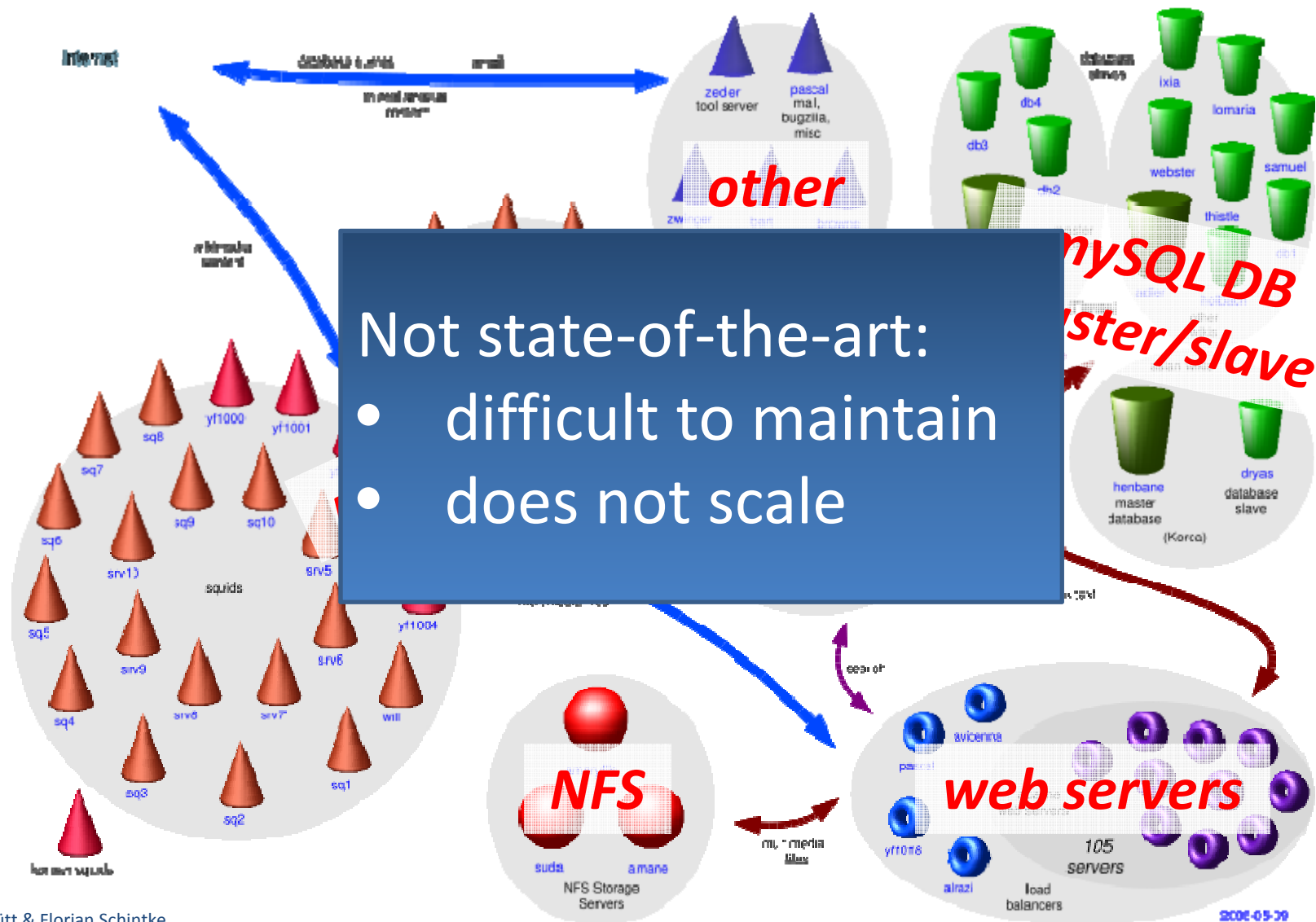
Zuse Institute Berlin

# Wikipedia: A top 10 Web site

**50.000 requests/sec**

- 95% answered by squid proxies  
→ **~18 squid servers**
- **2,000 req./sec** hit the backend  
→ **12 MySQL DB, ~158 Apache servers**

# The Wikipedia System Architecture



# Our Approach: P2P with Transaction Layer

## Benefits

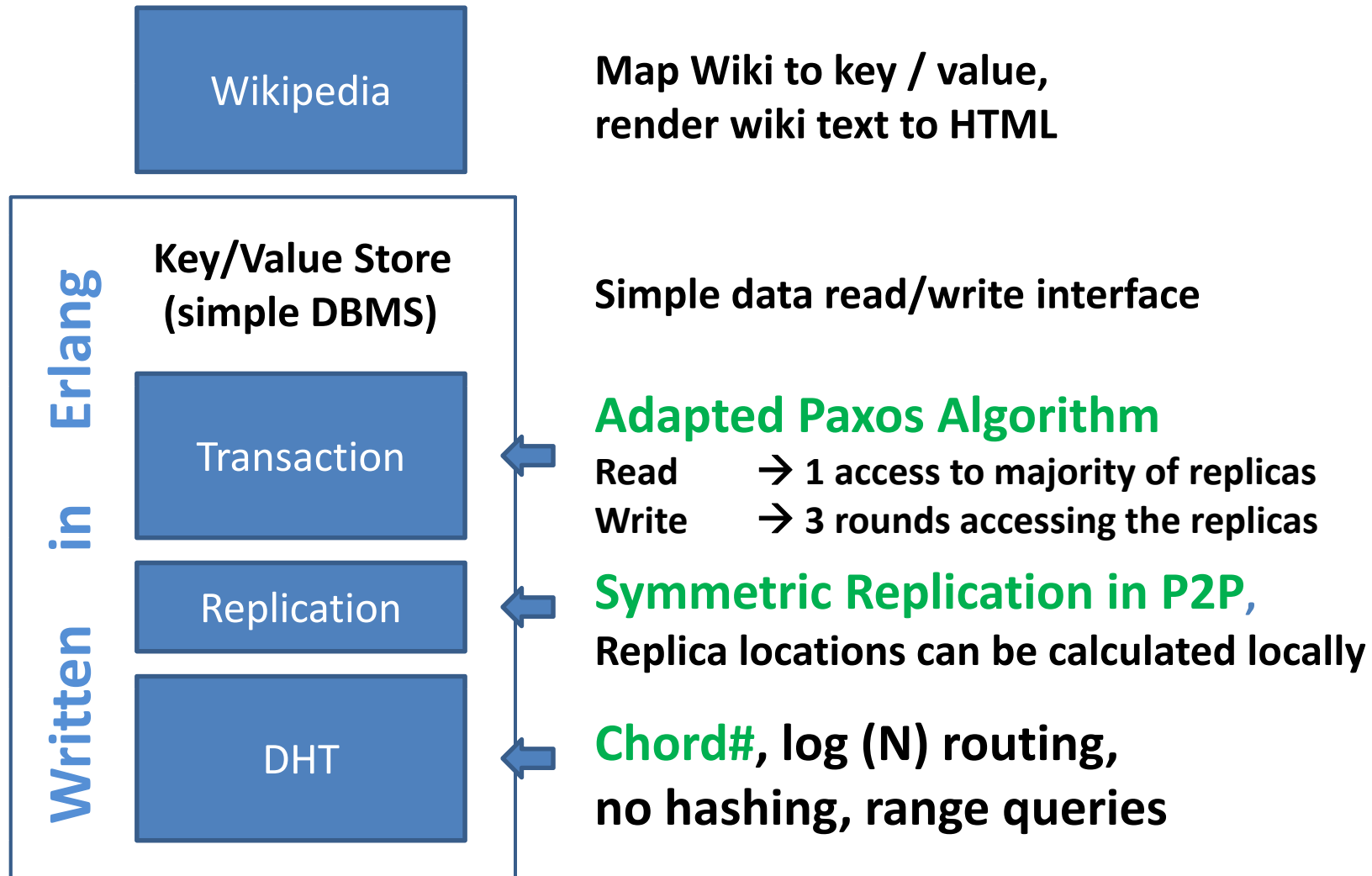
- distributed
- scalable
  - because of peer concept
- fault-tolerance
  - because of replication

## Challenges

- need synchronization
  - concurrency control
- need atomicity
  - in face of churn
- need transactions

DHT + Transactions = Scalable, Reliable, Efficient Key/Value Store

# System Solutions



# Demonstration

Two independent instances are setup:

## Cluster:

640 peers on 20 x 8 cores

## PlanetLab:

about 150 peers  
distributed worldwide



# Boot-Server: P2P management interface

- store keys
- search keys
- see the P2P ring
- statistics
- debug data



The screenshot shows the Chord# Boot Server Info Page. On the left is a navigation menu with the ZIB logo at the top. The menu items are: Home, Ring, Geo map, RRD Status, RRD Messages, Docu, Transactions, and Debug. The main content area on the right has a light blue background and contains the following sections:

- Chord# Boot Server Info Page**: Number of nodes: 161
- Simple Storage**: Add Key Value section with input fields for Key and Value, and Add and Reset buttons.
- Search**: Key input field and Lookup and Reset buttons.
- Args**: Last update: 16:50:25

© 2008 Zuse Institute Berlin - Computer Science Research, **Imprint**

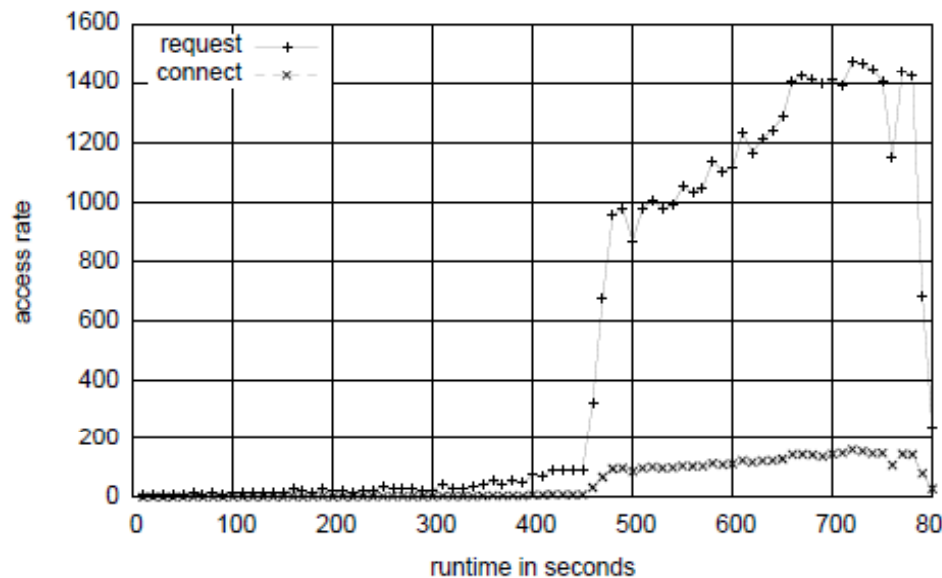
# Wikipedia Frontend

- Wikipedia on top of **scalable key/value store**
- installed a dump of **Simple English**
- interface language is static (Bavarian)
- **no images**  
URLs not in dump
- **browse links**
- **no fulltext search**

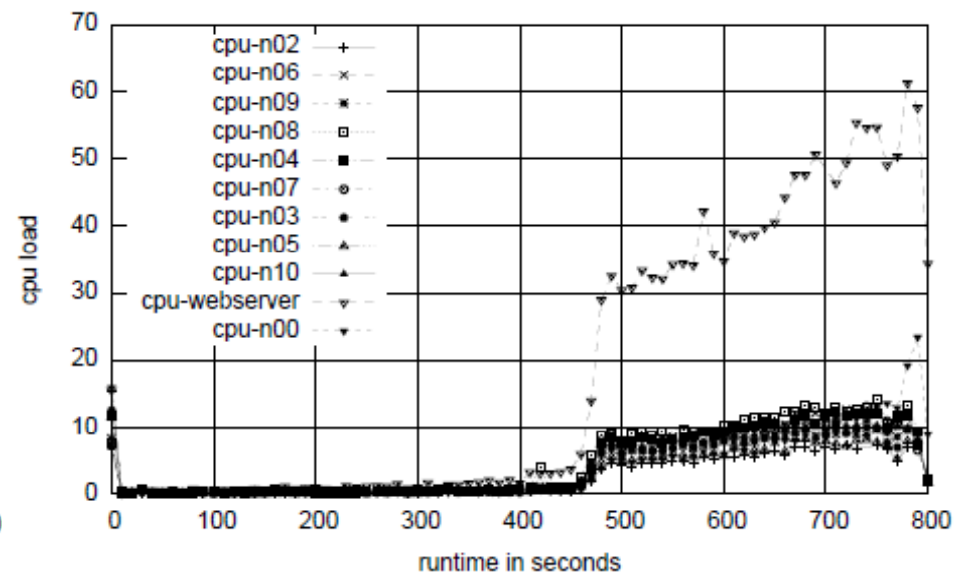


# Performance on Linux Cluster

1500 transactions/sec on 10 CPUs



throughput with increasing access rate over time



CPU load with increasing access rate over time

2500 trans./sec with 16 CPUs (64 cores) and 128 DHT nodes

# Performance

	Wikipedia	P2P-Wikipedia
Servers	<b>170</b>	<b>8</b>
Requests / s	<b>2000</b>	<b>2500</b>

# Demonstration

Everybody can participate now:

**[www.onscale.de/scale.html](http://www.onscale.de/scale.html)**

load & modify pages

# Detailed Descriptions

## Wiki

S. Plantikow, A. Reinefeld, F. Schintke.

**Transactions for Distributed Wikis on Structured Overlays.**

DSOM 2007.

## Transactions

M. Moser, S. Haridi .

**Atomic Commitment in Transactional DHTs.**

1st CoreGRID Symposium, Aug 2007.

## DHT

T. Schütt, F. Schintke, A. Reinefeld.

**A Structured Overlay for Multi-dimensional Range Queries.**

Euro-Par 2007.

T. Schütt, F. Schintke, A. Reinefeld .

**Structured Overlay without Consistent Hashing: Empirical Results.**

GP2PC 2006.

# Summary

DHT + Transactions = Scalable, Reliable, Efficient Key/Value Store

- Previously, P2P was mainly used for file sharing (read only).
- **We support consistent, distributed *write* operations.**
- Numerous applications
  - Web 2.0, internet databases, transactional online-services, ...

# System Layers

