A Scalable Content-Addressable Network [1]

S. Ratnasamy, P. Francis, M. Handley, R. Karp and S. Shenker

Proseminar “Algorithmen für Rechnernetze”

Lehrstuhl für Rechnernetze und Telematik,

Prof. Dr. Christian Schindelhauer

Lukas Veneziano

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Motivation

- Regular client-server structures

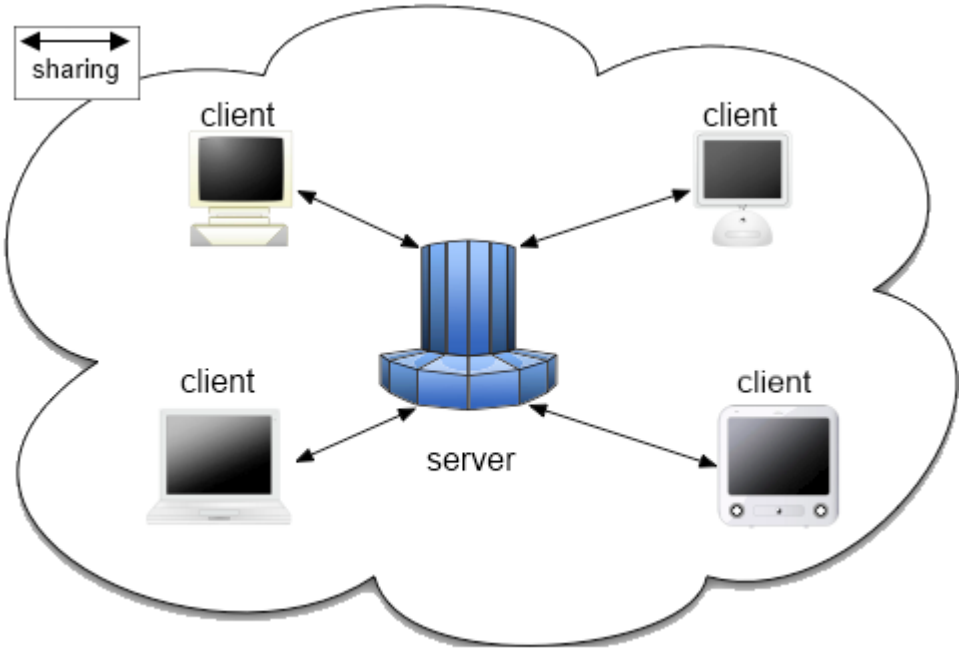
Pros

- Easy to use for the user
- Easy to maintain

Cons

- High traffic
- Low fault tolerance
  - Single point of failure
- Scalability

➔ High costs



Client-server communication [Graphic 1]

Motivation

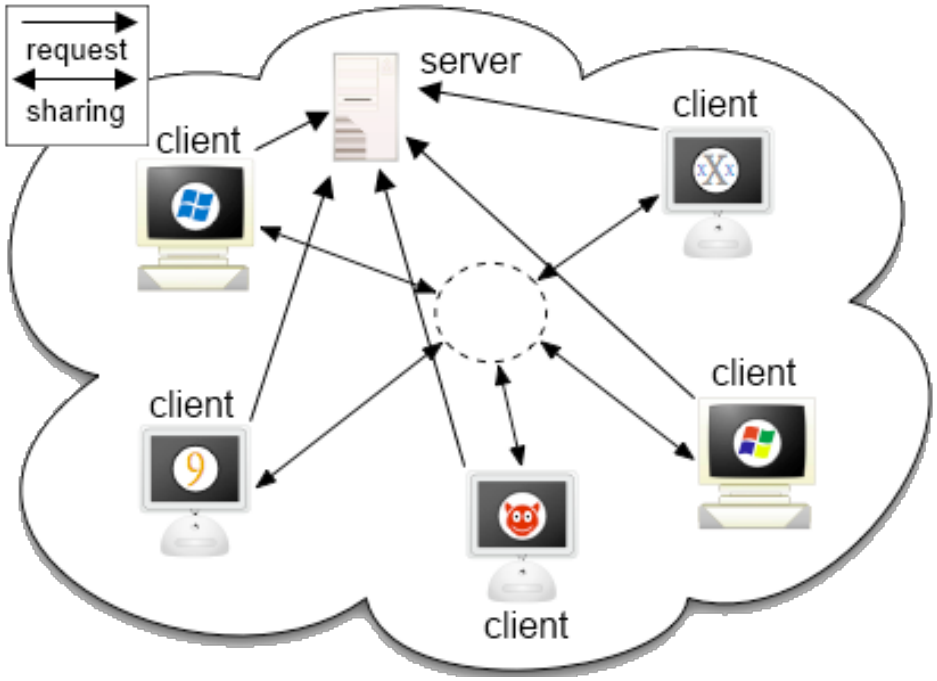
- Regular peer-to-peer communication

Pros

- Low traffic for server
- High speed for peer
- Better fault tolerance

Cons

- Find other peers
- Server needed



Peer-to-Peer communication [Graphic 2]

Motivation

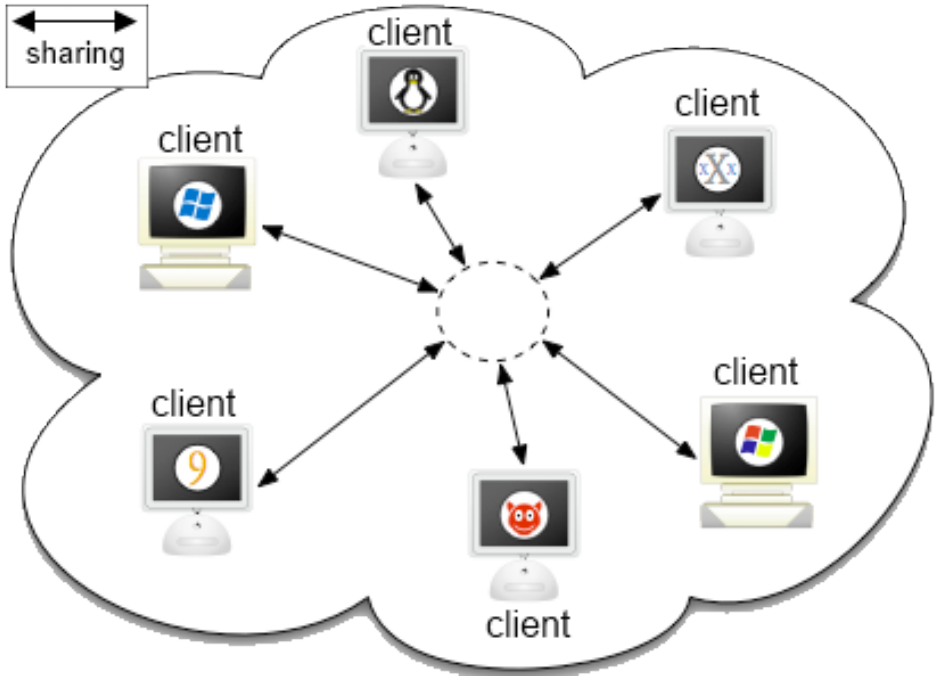
- Distributed hash table network (DHT)

Pros

- Low traffic for server
- High speed for peer
- High fault tolerance
- Self-organizing

Cons

- Mostly Usersoftware

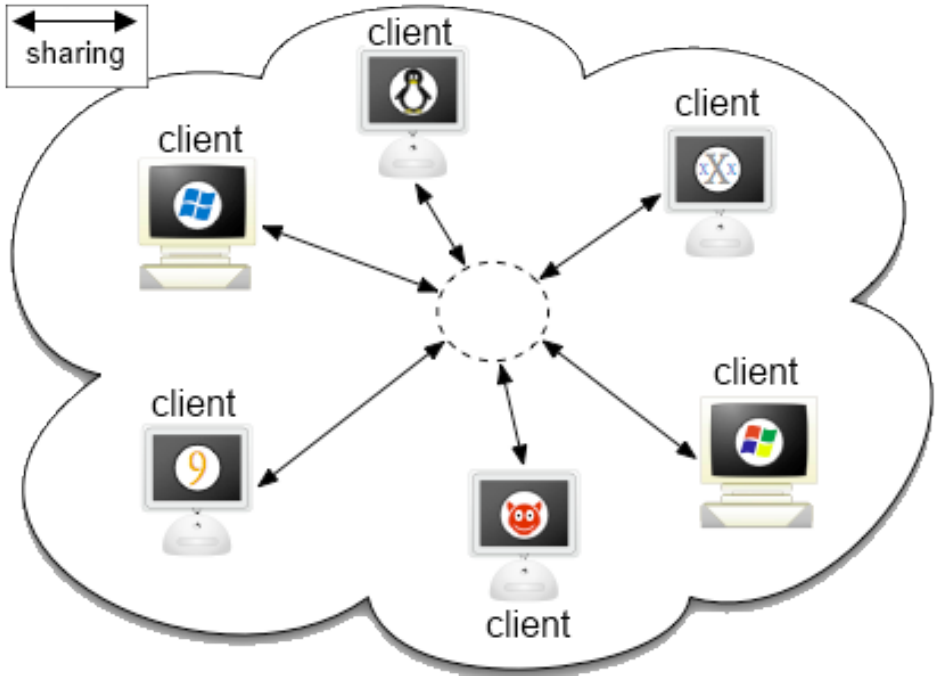


Peer-to-peer with DHT communication [Graphic 1]

Motivation

- Distributed hash table network (DHT)

  - Four DHT proposals (CAN, Chord, Pastry, Tapestry)
  - Unlike Napster etc.
    - Usable as back-end
    - Search for Data in the network
    - → No server needed

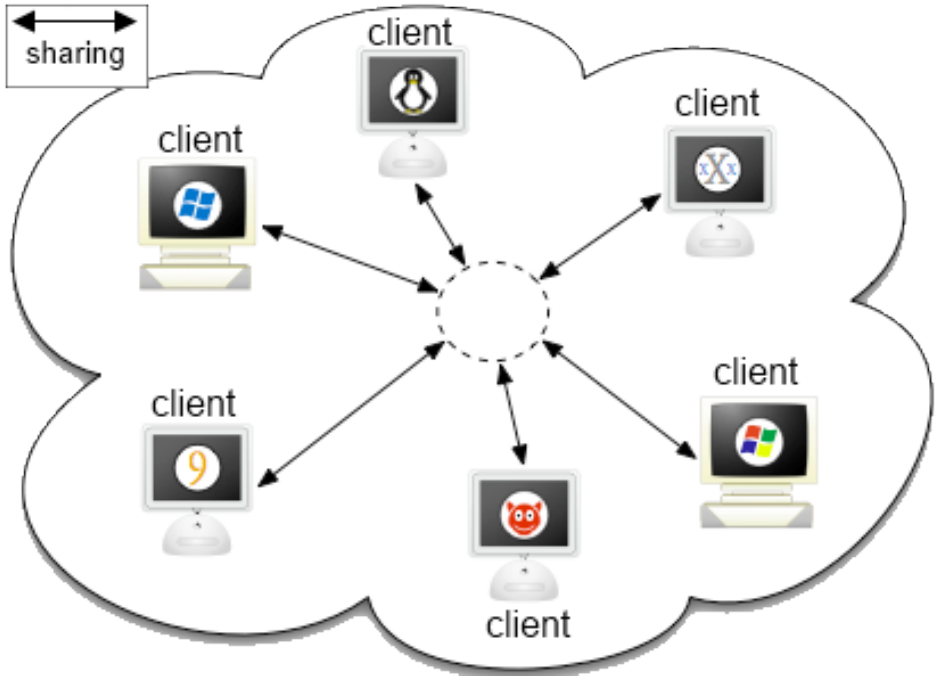


Peer-to-peer with DHT communication [Graphic 1]

Motivation

- Content-Addressable Network (CAN)

- Distributed, decentralized P2P infrastructure
- High scalability
- High fault tolerance
- Easy to search (hash-table)



Peer-to-peer with DHT communication [Graphic 1]

CoNe Freiburg

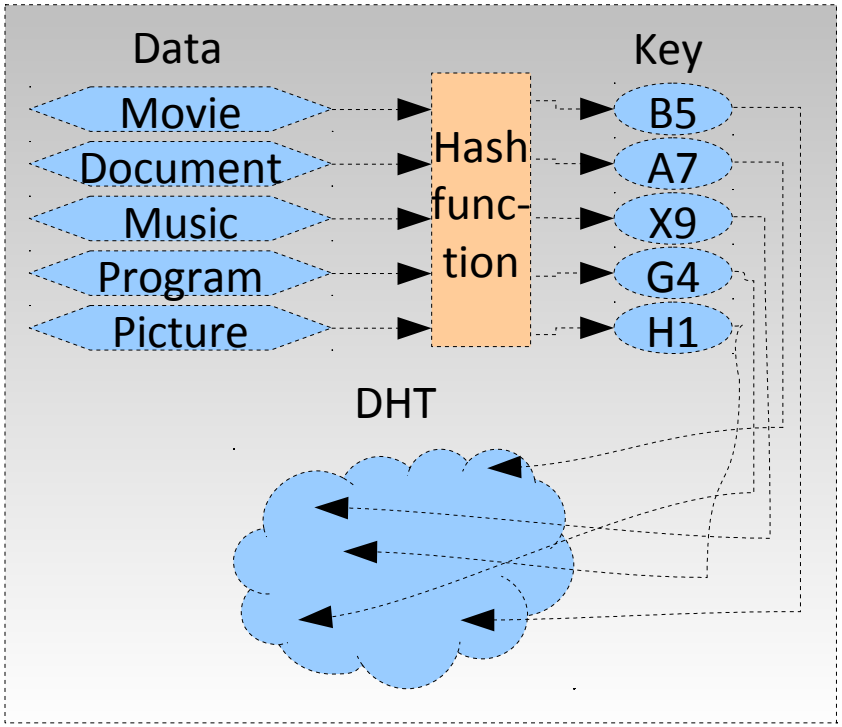
Overview

1. **Design**
   - Routing
   - Construction
   - Maintenance
2. Improvements
3. Measurement data

UNI FREIBURG

Design

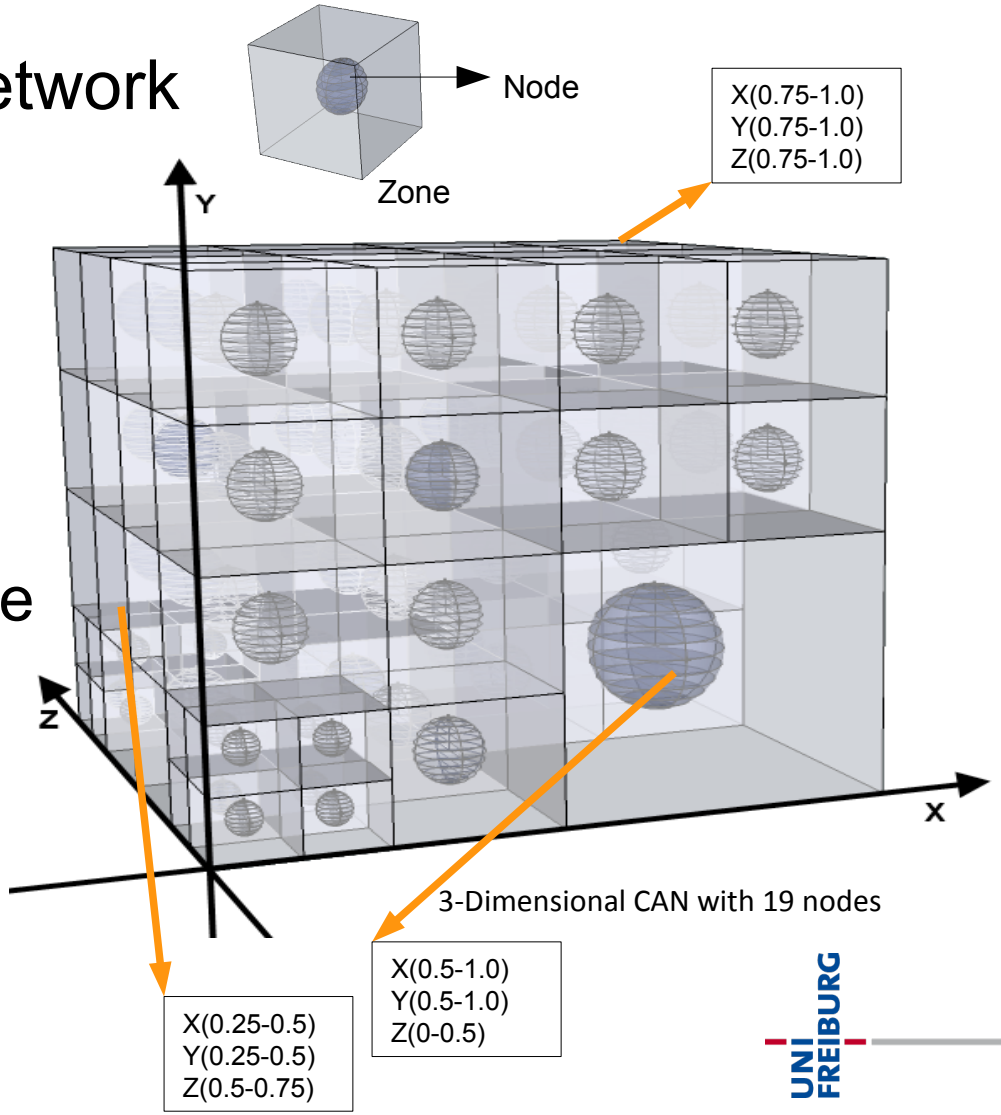
- Content Addressable Network
  - Mapped to area of $[0,1]^n$
  - Organized with hash-table



Concept of a hash-algorithm

Design

- Content Addressable Network
  - Data is stored in zones
    - Each node represents one zone
  - N-dimensional **torus** (not visible in graphic)
  - Paths are short because of torus structure

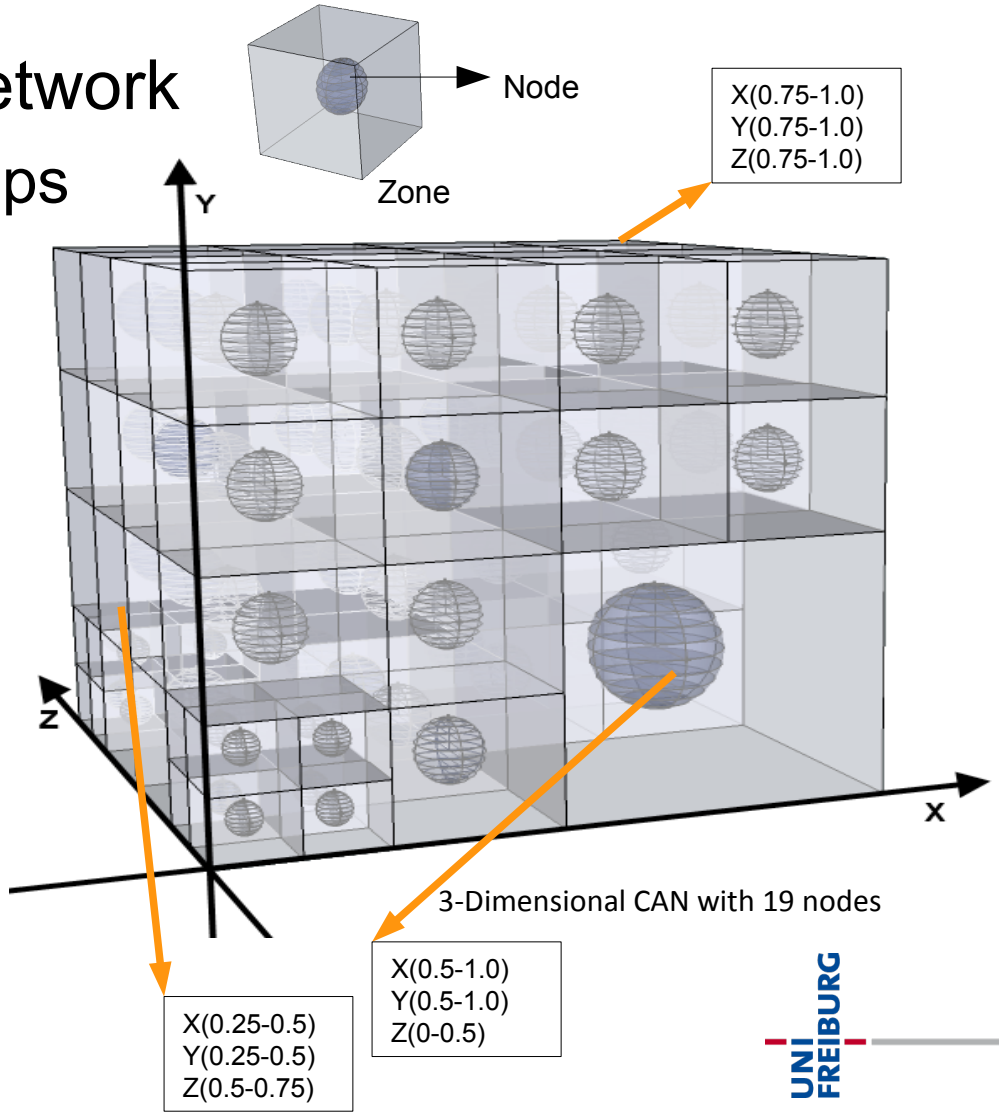


Design

- Content Addressable Network
  - Expected number of hops for CAN in d-dimensions:

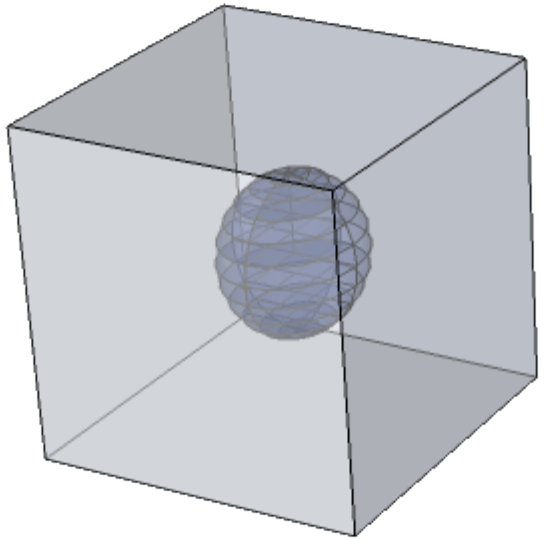
    $O(\sqrt[d]{n})$
  - Average degree of a node:

    O(d)



Design

- One node
  - Knows its zone
  - Knows its neighbours
  - Knows the specification of the specific CAN
  - ➔ Can act autonomic



One node of a CAN

Formula for hops and degree from [2]

CoNe Freiburg

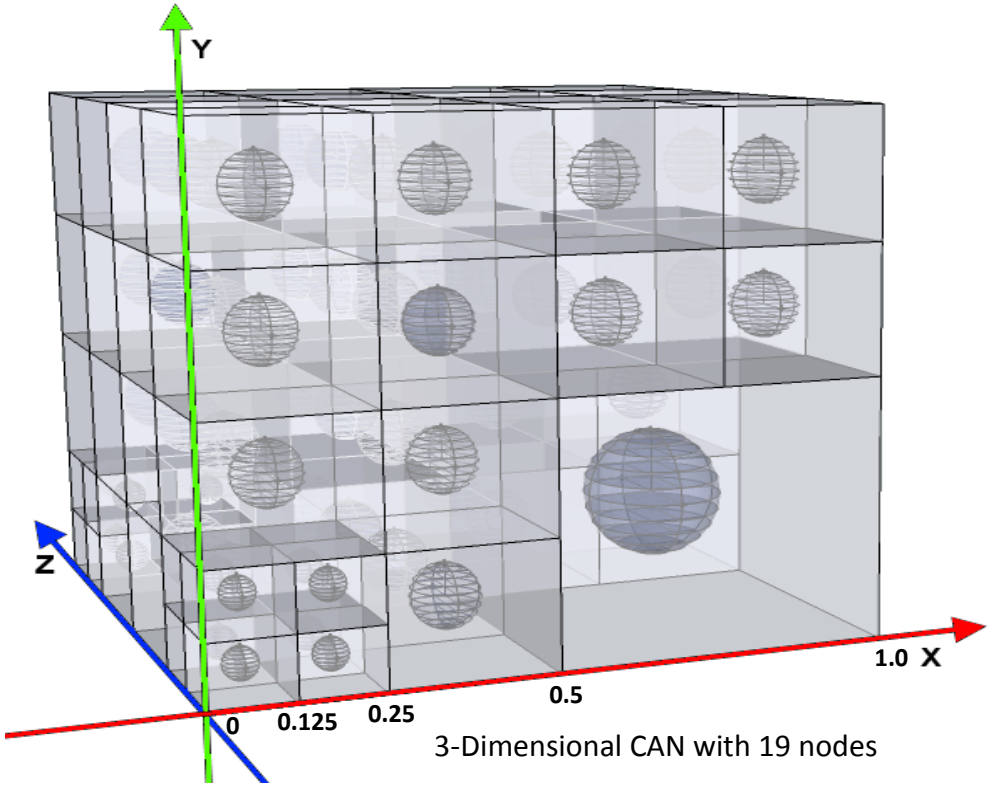
Overview

1. Design
   - **Routing**
   - Construction
   - Maintenance
2. Improvements
3. Measurement data

UNI FREIBURG

Routing

- Simple routing along the axes
  - e.g.
    1. X-axis
    2. Y-axis
    3. Z-axis



3-Dimensional CAN with 19 nodes

CoNe Freiburg

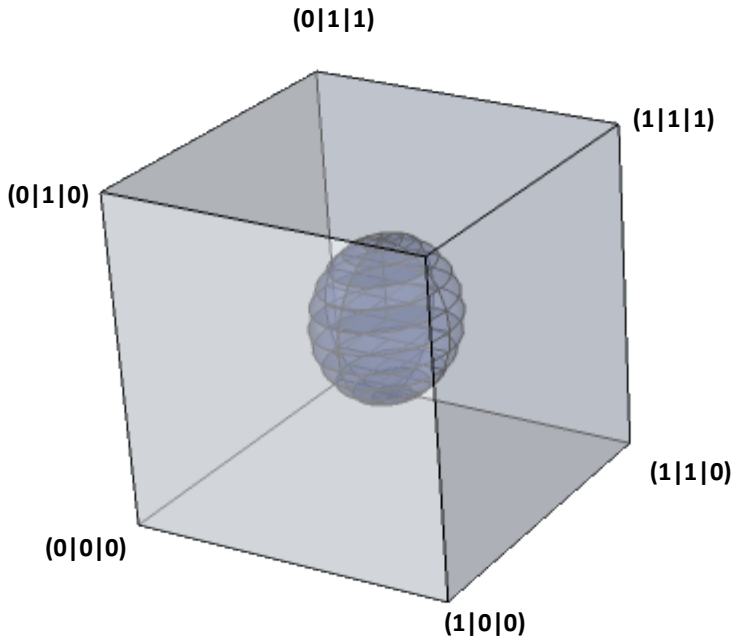
Overview

1. Design
   - Routing
   - **Construction**
   - Maintenance
2. Improvements
3. Measurement data

UNI FREIBURG

Construction

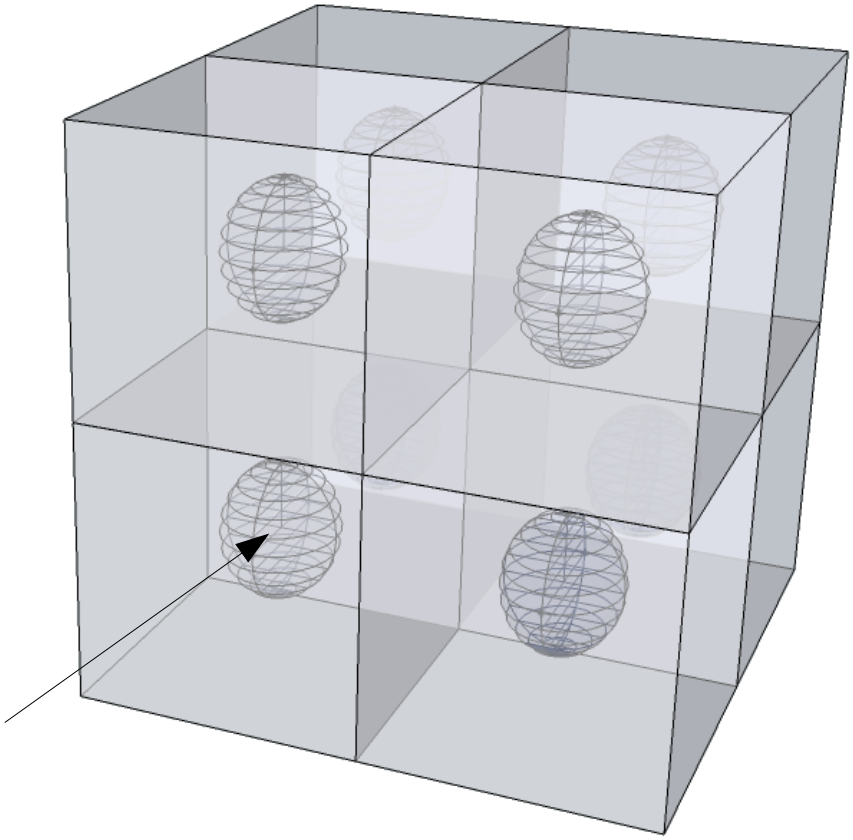
- One node is a CAN
- Until other nodes are added



A CAN which consists of only one node

Construction

- Add other nodes
  1. Find a node in the CAN

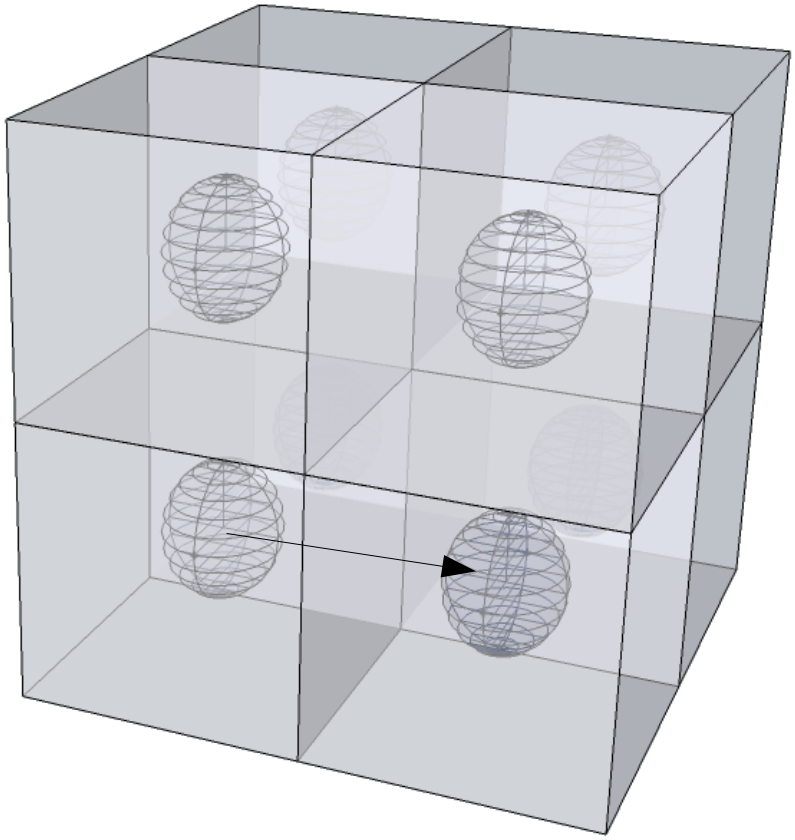


A CAN of 8 nodes

CoNe Freiburg

Construction

- One node can be a CAN
- Add other nodes
  1. Find a node in the CAN
  2. Route to a random zone



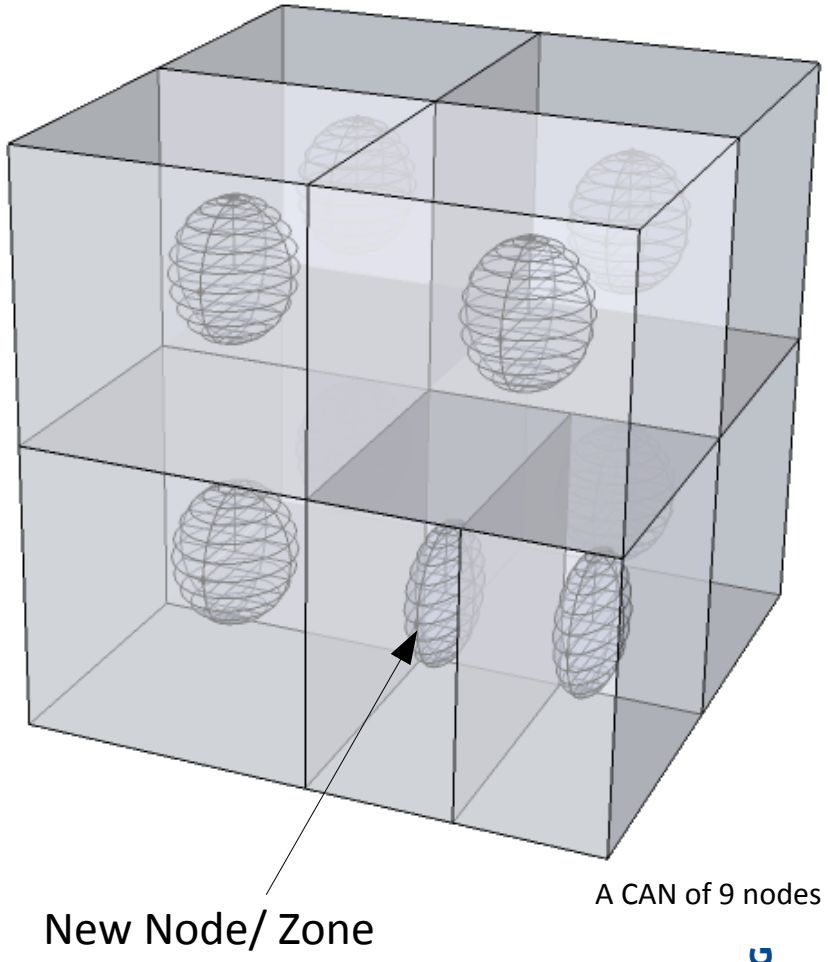
A CAN of 8 nodes

UNI FREIBURG

CoNe Freiburg

Construction

- One node can be a CAN
- Add other nodes
  1. Find a node in the CAN
  2. Route to a random zone
     a. Split the random zone

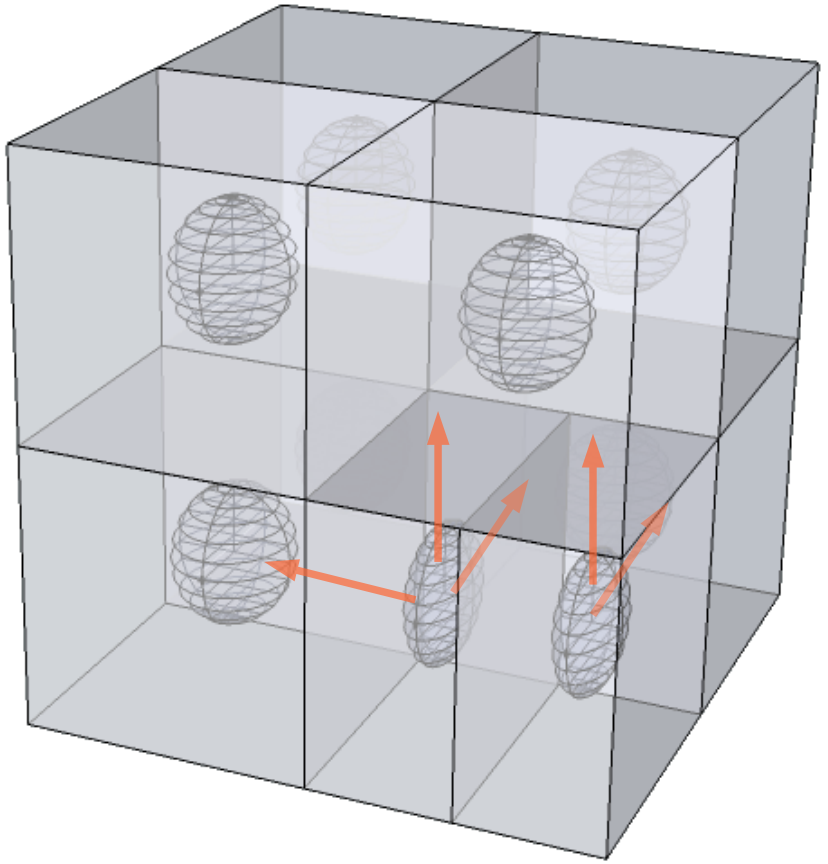


UNI FREIBURG

CoNe Freiburg

Construction

- One node can be a CAN
- Add other nodes
  1. Find a node in the CAN
  2. Route to a random zone
     a. Split the random zone
  3. Inform the neighbours



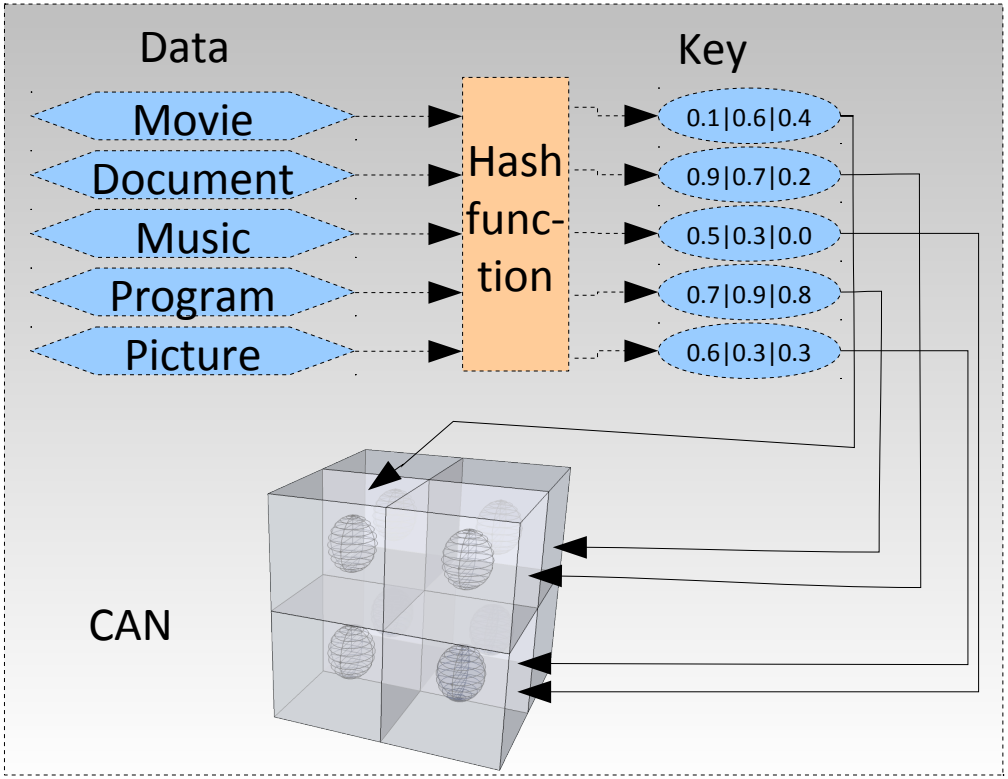
A CAN of 9 nodes

UNI FREIBURG

Construction

Store Data in a CAN

- Find a node
- Transfer data into key
- Route data by key to zone



CoNe Freiburg

Overview

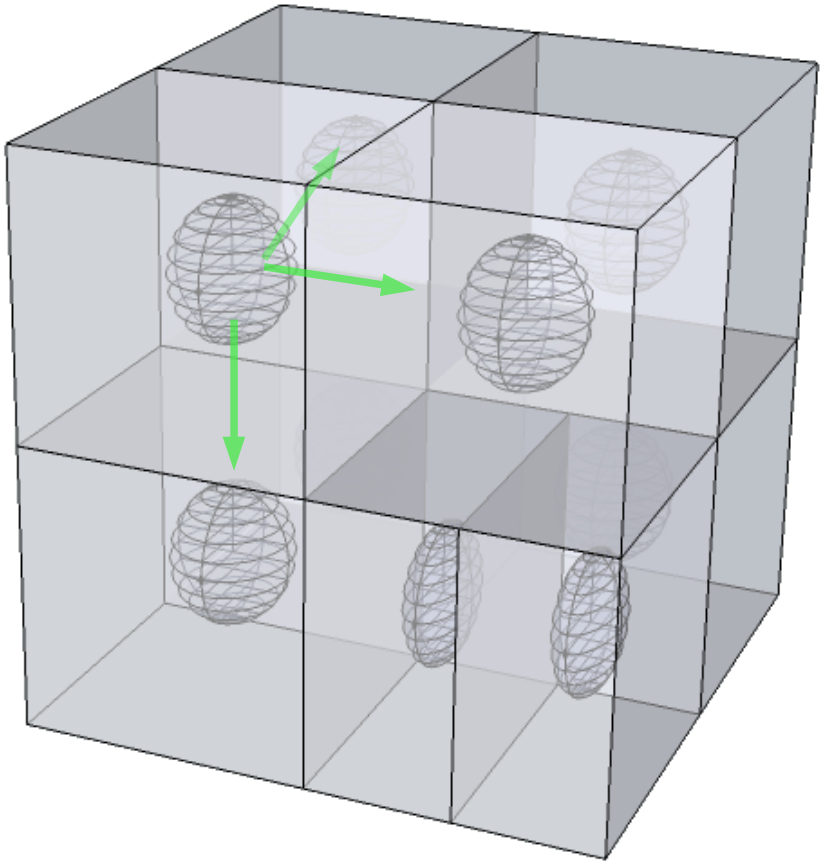
1. Design
   - Routing
   - Construction
   - **Maintenance**
2. Improvements
3. Measurement data

UNI FREIBURG

CoNe Freiburg

Maintenance

- Each node sends alive messages to its neighbours

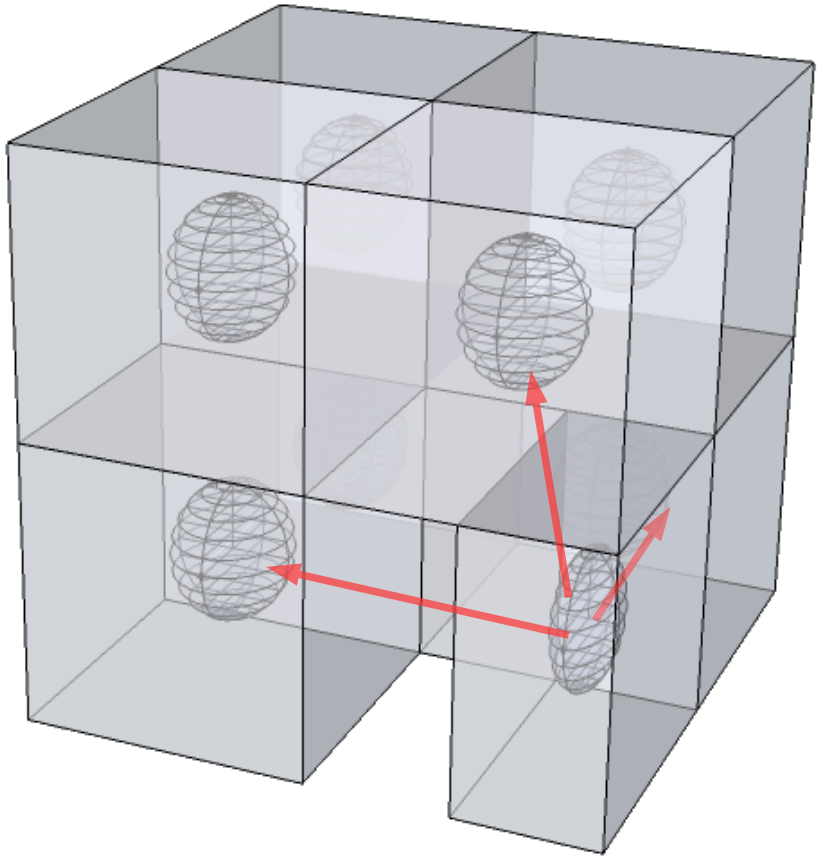


A CAN of 8 nodes

UNI FREIBURG

Maintenance

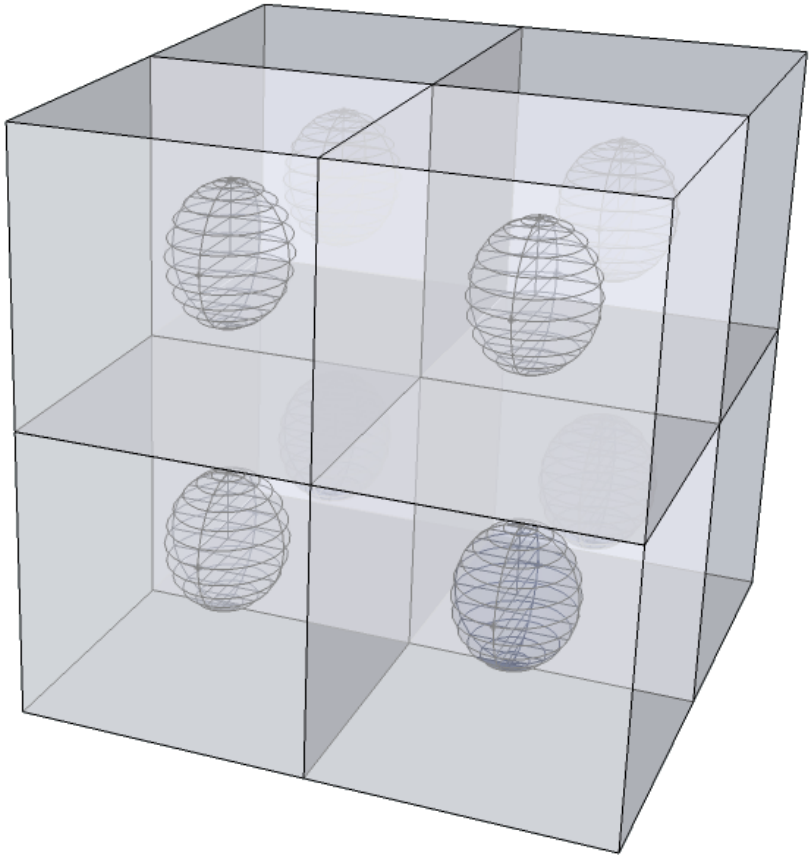
- Each node sends alive messages to its neighbours
- Takeover of dead nodes
  - Takeover message



A CAN of 8 nodes

Maintenance

- Each node sends alive messages to its neighbours
- Takeover of dead nodes
  - Takeover message



A CAN of 8 nodes

CoNe Freiburg

Overview

1. Design
   - Routing
   - Construction
   - Maintenance
2. **Improvements**
3. Measurement data

UNI FREIBURG

CoNe Freiburg

Improvements

- Many improvements for speed and fault tolerance for the CAN-concept
  - More dimensions
  - Redundant coordinate-spaces
  - Routing with IP-topology
  - Node-allocation with IP-topology
  - Node-allocation with information of scope
  - Zone is managed by multiple nodes
  - Multiple hash-algorithms
  - Caching

[Graphic 4]

UNI FREIBURG

CoNe Freiburg

Overview

1. Design
   - Routing
   - Construction
   - Maintenance
2. Improvements
3. **Measurement data**

UNI FREIBURG

Measurement data

- Used metrics
  - Path length
  - Neighbour Count
  - Latency
  - Volume
  - Fault tolerance
  - Hash table availability

[Graphic 5]

CoNe Freiburg

Measurement data - parameters

Parameter	“bare bones” CAN	“knobs on full” CAN
dimensions	2	10
realities	1	1
peers per zone	0	4
# hash functions	1	1
RTT weighted routing metric	OFF	ON
uniform partitioning	OFF	ON
landmark ordering	OFF	OFF

data from [1]

UNI FREIBURG

CoNe Freiburg

Measurement data - results

Metric	“bare bones” CAN	“knobs on full” CAN
Path length	198.0	5.0
# neighbours	4.57	27.1
# peers	0	2.95
IP Latency	115.9ms	82.4ms
CAN Latency	23,008ms	135.29ms

data from [1]

UNI FREIBURG

CoNe Freiburg

Conclusion

- powerful technique to store data in an internet-like scale
- much more scalable, fault tolerant and faster than a regular server-structure
- with some improvements even more powerful

UNI FREIBURG

Thank you for your attention

CoNe Freiburg

References

- [1] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A Scalable Content-Addressable Network. ACM, San Diego, California, USA, 2001
- [2] C.Schindelhauer (2010), Peer to Peer Networks, Slides from Lecture Peer to Peer Networks (2010) by Prof. Dr. C. Schindelhauer http://archive.cone.informatik.uni-freiburg.de/teaching/vorlesung/peer-to-peer-s10/
- [Graphic 1], [Graphic 2], [Graphic 3] http://www.semanticmedia-showcase.de/MuseumSM/MMORPGS/Peer2Peer.htm , abgerufen 15.07.2012, 15:00 Uhr und bearbeitet.
- [Graphic 4] http://2.imimg.com/data2/AC/TG/MY-2100409/improve-your-bottomline-250x250.jpg , abgerufen 16.07.2012, 19:00 Uhr
- [Graphic 5] http://4.bp.blogspot.com/jOevxvE1UTY/TrxNVprRfVI/AAAAAAAAAjU/FexOBWk8nUg/s1600/measurement_ruler.jpg , abgerufen 16.07.2012, 19:00 Uhr

UNI FREIBURG