

Self-organizing Distributed Data Access

Karl Aberer
School of Computer and Communication Sciences
EPFL, Switzerland

karl.aberer@epfl.ch
lsirwww.epfl.ch
www.p-grid.org

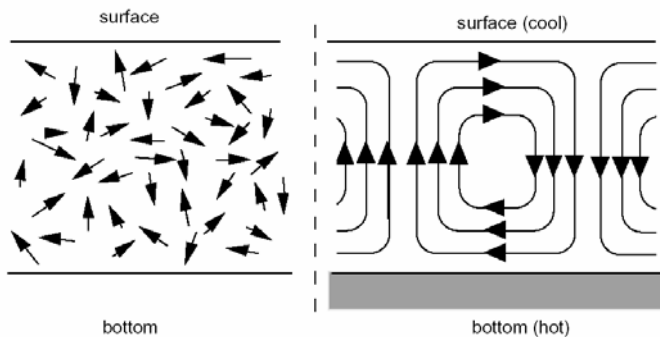
GI Arbeitsgespräch
"Qualität in Peer-to-Peer Netzen", 24.9.2003, Darmstadt.

joint work with:
Anwitaman Datta, Manfred Hauswirth, Roman Schmidt

Overview

1. Self-organization and distributed data access
|
2. Scalable distributed data access
|
- 3. P-Grid: scalable, self-organizing distributed data access
|
4. Updates in a P2P environment
|
- 5. Self-healing directories: P-Grid and P2P updates in action
|
- 6. Conclusion

1. What is Self-organization ?

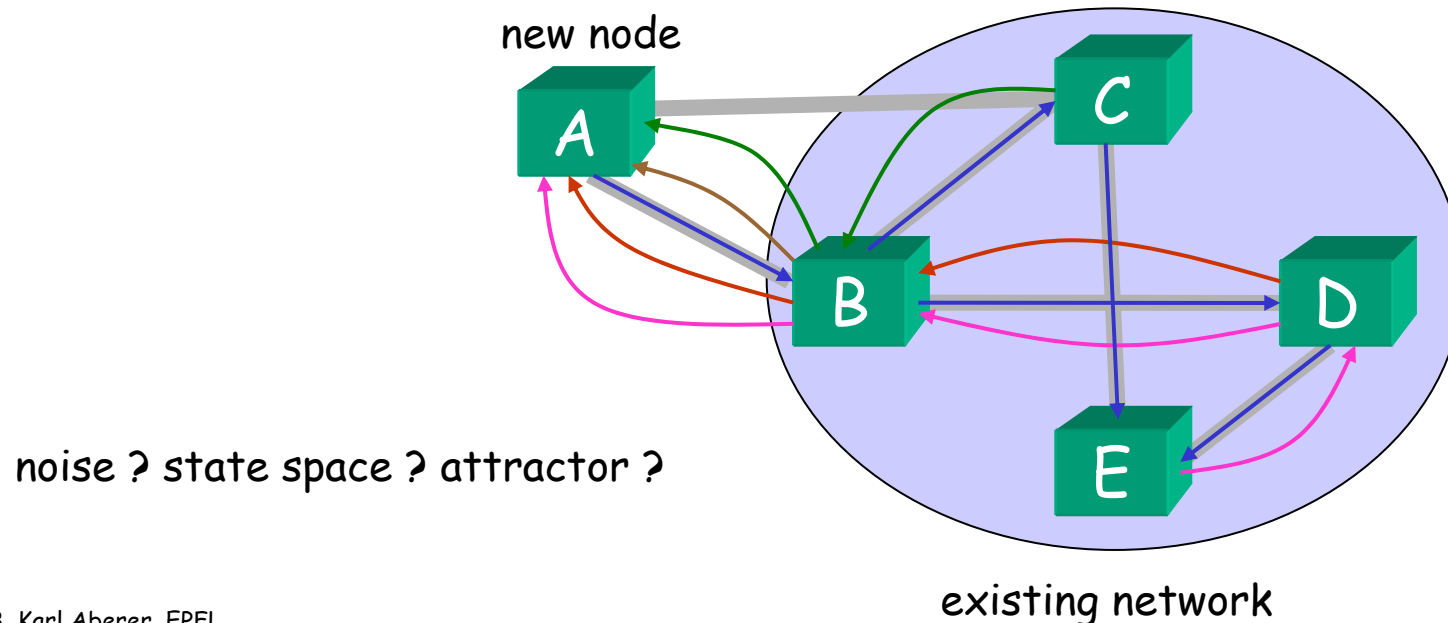


- Informal characterization (physics, biology, cybernetics,... and CS)
 - distribution of control (= decentralization = symmetry in roles = P2P)
 - local interactions, information and decisions (= autonomy)
 - emergence of global structures
 - failure resilience and scalability
- Formal characterization (Francis Heylighen)

"The basic mechanism underlying self-organization is the **noise**-driven variation which explores different regions in a system's **state space** until it enters an **attractor**"

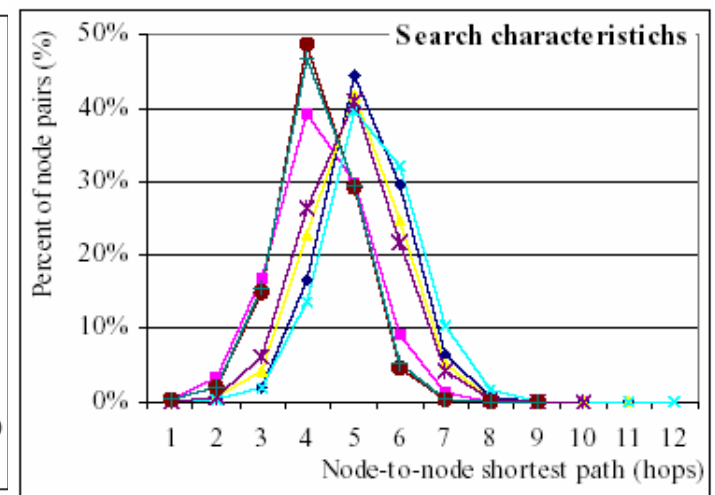
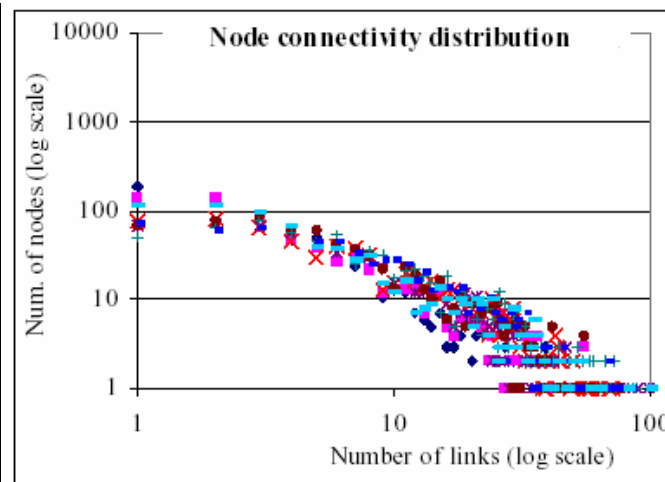
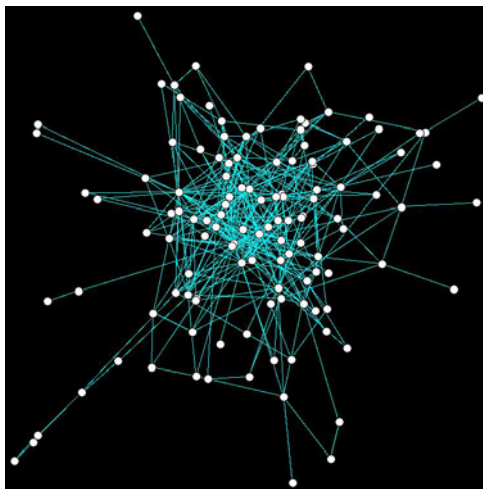
Self-organizing Data Access

- The most studied system: *Gnutella* and descendants
 - A **group membership protocol** (ping/pong) to construct an **overlay network**
 - A **search protocol** (query/queryhit) to locate information in the overlay network
- What is *Gnutella* used for ?
 - content sharing (music, video, etc.)
 - studying self-organization in computer science
- Why is *Gnutella* self-organizing ?



Resulting Network Structure

- Specific network structure emerges from a self-organizing process
 - state space = graphs with constant outdegree
 - noise = time of node joins, latency, decisions on connectivity
 - attractor = approx. powerlaw graphs
- Powerlaw graphs
 - Preferential attachment + growing network [Barabasi, 1999]
 - However uniform distribution at tail (minimal connectivity, adds to robustness)
 - Short diameter [Ripenau et al, 2002]



Searching the Network

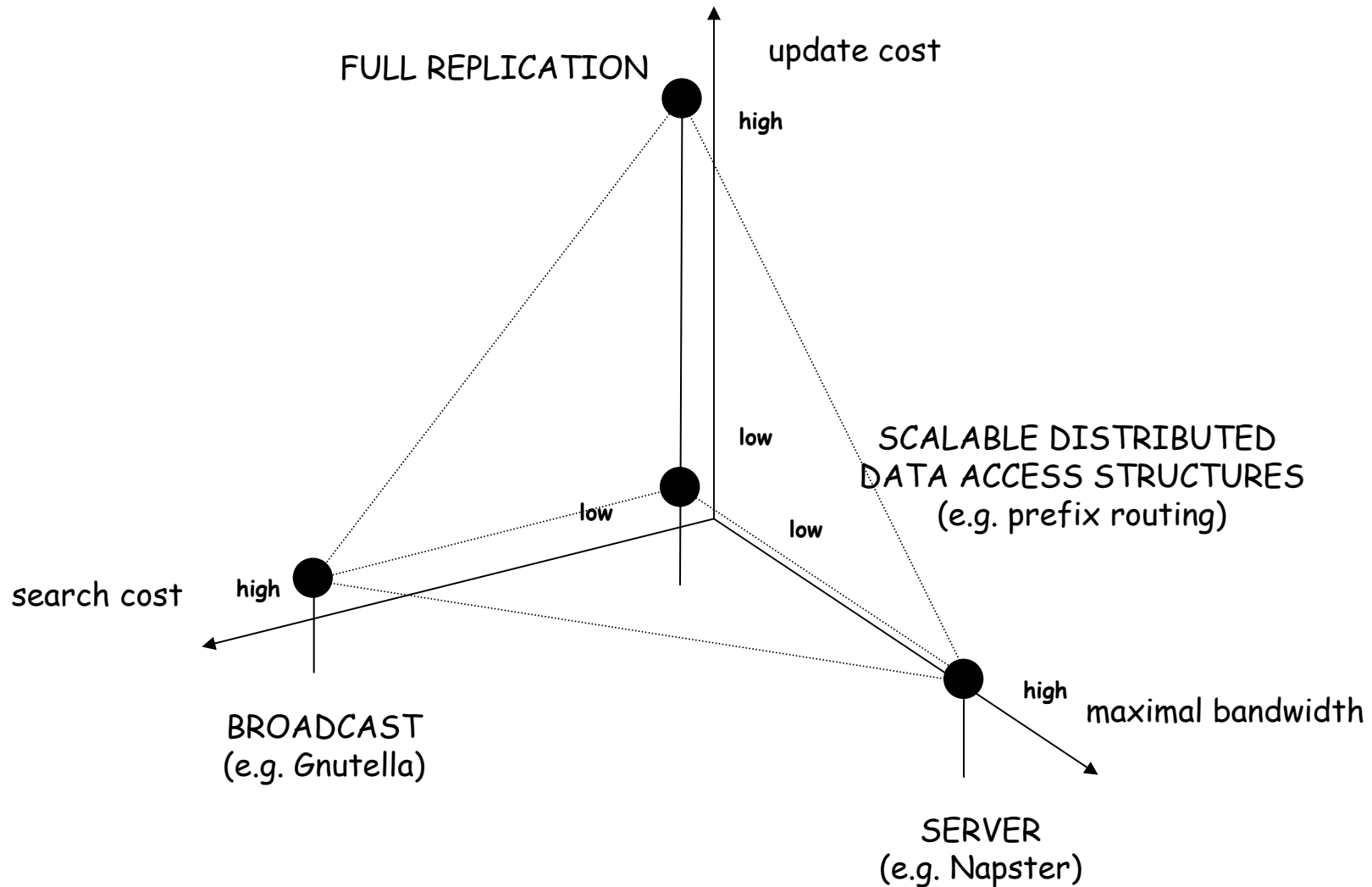
- Gnutella = unstructured network
 - = no information about what data neighboring nodes store
 - Broadcast to d neighbors with maximal time-to-live TTL
 - $TTL >$ network diameter
 - duplicate message detection: $< dN$ messages
- Reducing message bandwidth
 - Random walkers [Lv et al, 2002]
 - Percolation search [Sarshar et al 2002]
 - Assumptions on network structure (random, powerlaw, small world graph etc.)

Conclusion: Network construction and search protocols are orthogonal issues

1. Network construction is a self-organizing process
2. Search is "simply" a randomized, distributed algorithm

Problem: Search in "unstructured networks" requires high bandwidth

2. Efficient Distributed Data Access



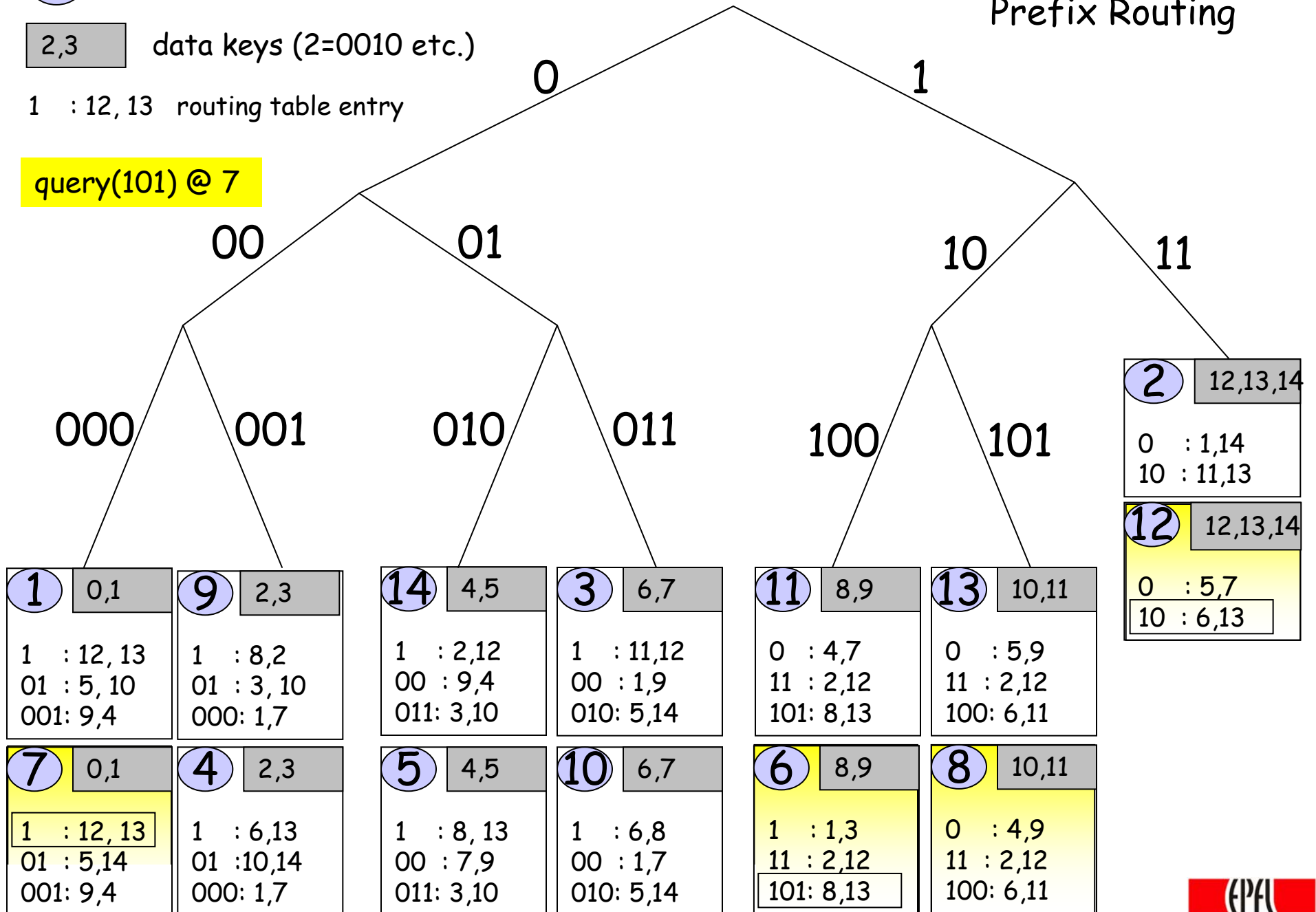
ID peer identifier

2,3 data keys (2=0010 etc.)

1 : 12,13 routing table entry

query(101) @ 7

Prefix Routing



1. Self-organization and distributed data access

2. Scalable distributed data access

3. Construction of a Prefix Routing Structure

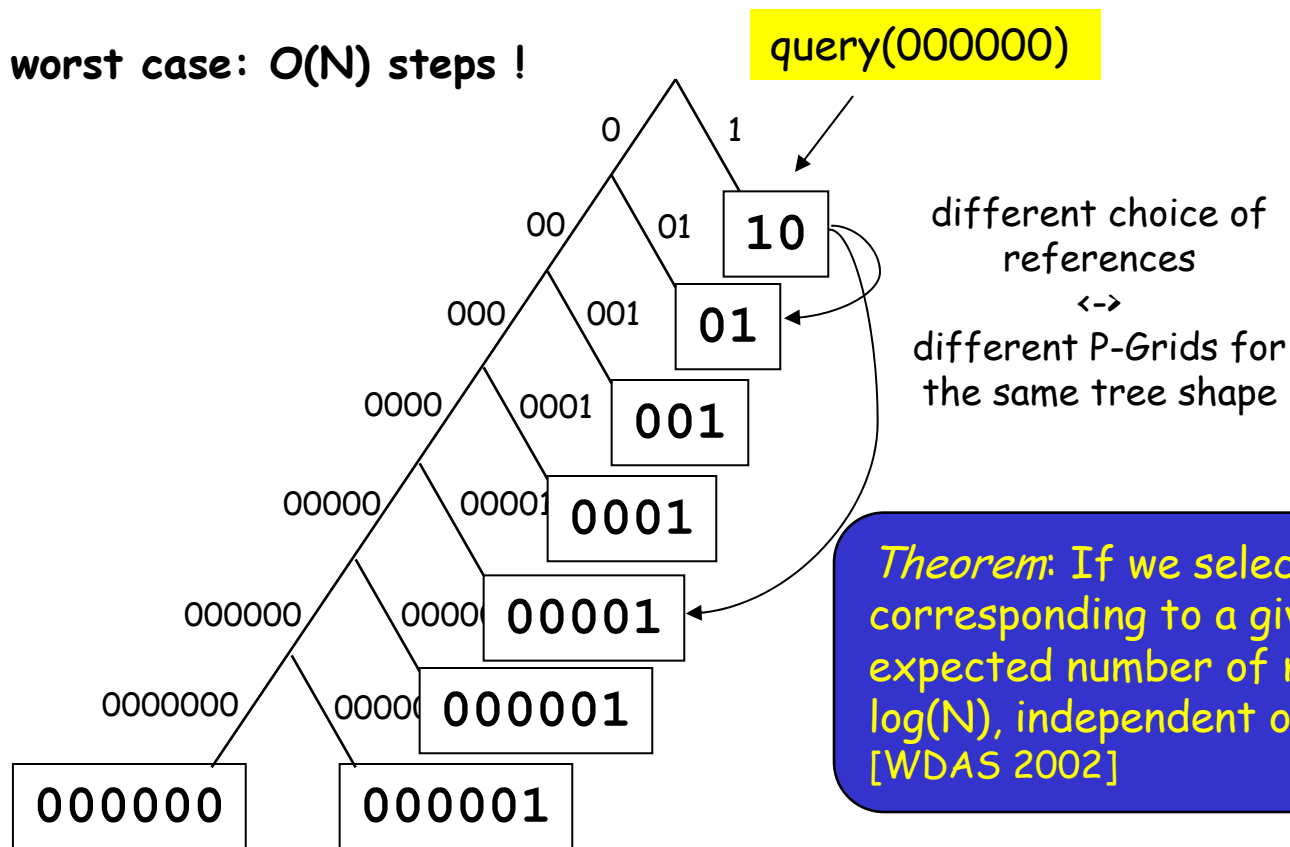
- Standard method (Plaxton, Pastry, Chord, Tapestry, ...)
 - static assignment of peers to data keys:
fully defines the distributed data structure
 - distributed, deterministic algorithms insert nodes correctly into the network
 - Problems: load balancing, robustness
- Self-organizing method (P-Grid [CoopIs 2001, IEEE IC 2002,...])
 - peers agree via local interactions on their data key
 - decide dynamically whether a data region is densely populated and requires further refinement of search tree

when two peers meet
peer extends path if #data items > threshold
do the necessary bookkeeping
otherwise data exchange (duplicate generation)

Unbalanced Search Trees

- Construct the tree for the following data:
10, 01, 001, 0001, 00001, 000001, 000000

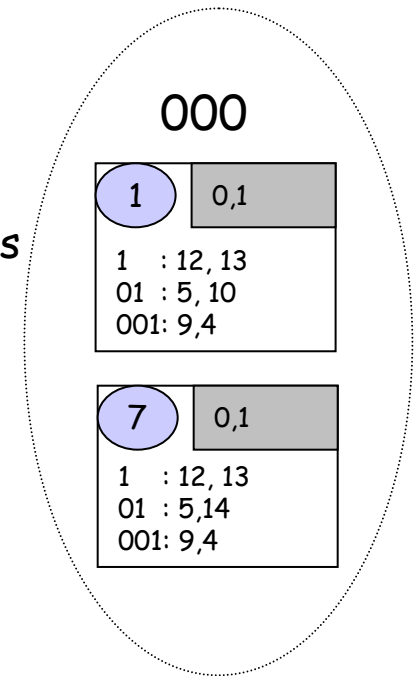
worst case: $O(N)$ steps !



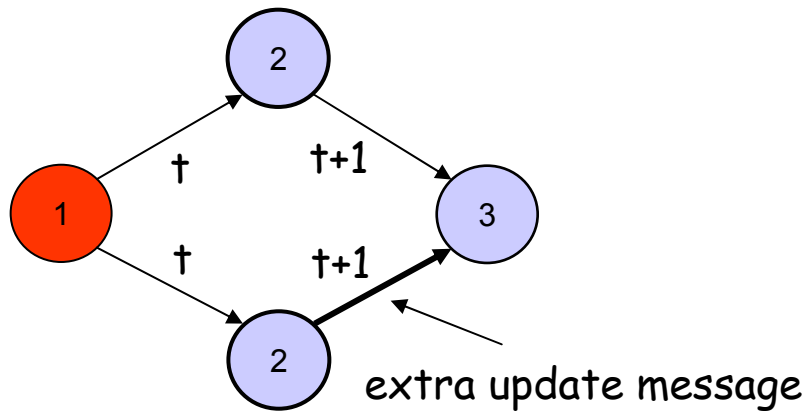
- P-Grid construction is a self-organizing process
- Prefix search is an (efficient) randomized, distributed algorithm

4. Updates

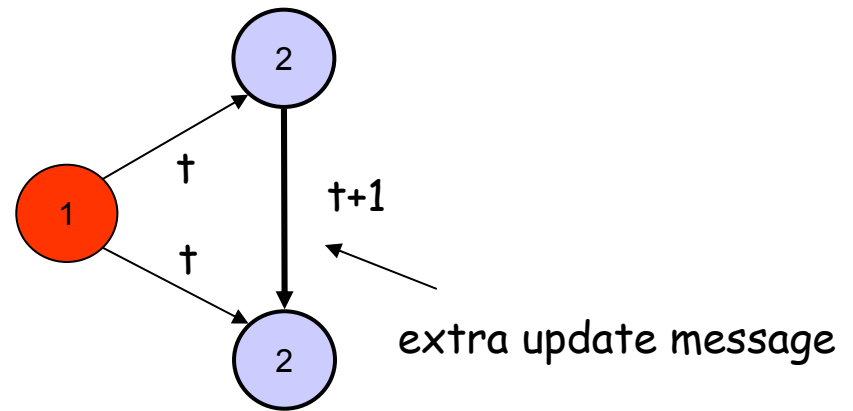
- Updates in the presence of replicas
 - There is no real alternative to flooding the update to all replicas
 - replicas form a P2P subnetwork
- Particular environment
 - availability of nodes is generally low (e.g. 10%)
 - eventual consistency is sufficient
 - communication overhead is critical
 - degree of replication is high
 - connectivity among replica nodes is high
- Basic approach [ICDCS 2003]
 - replicas going online request updates from other replicas (pull)
 - replicas push updates selectively (push)



Selective Push



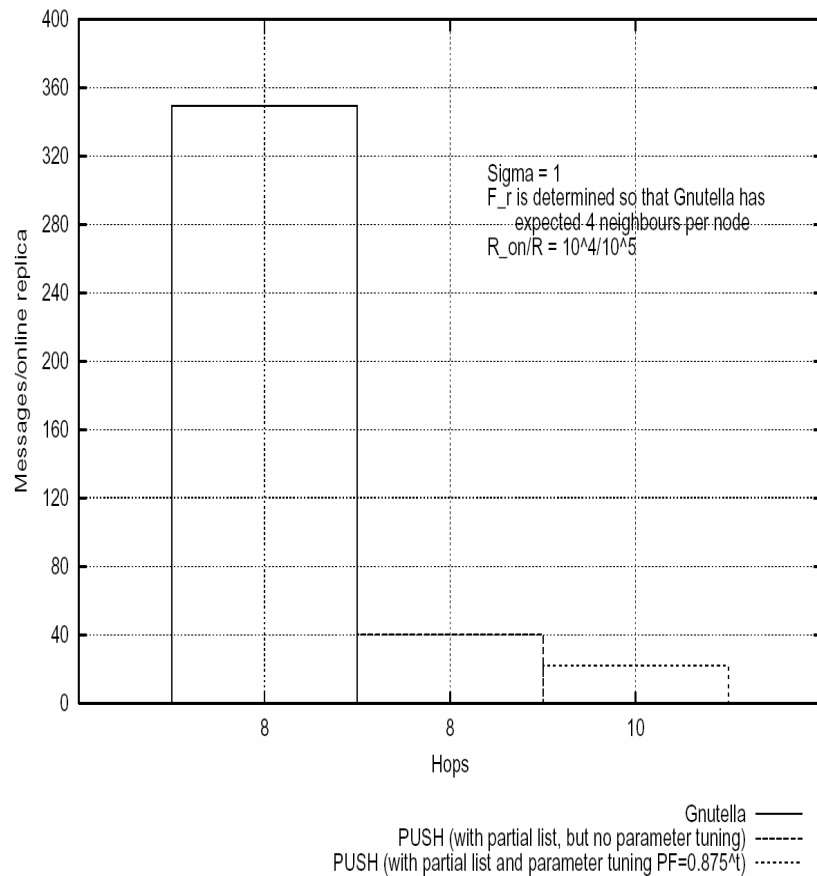
avoid parallel redundant update:
 messages are propagated only
 with probability $PF < 1$ and to
 a fraction of the neighbors



avoid sequential redundant update:
 partial lists of informed neighbors are
 transmitted with the message

Comparison to Other Gossiping Schemes

Results from analytical model (10^5 online peers, 10^4 online)



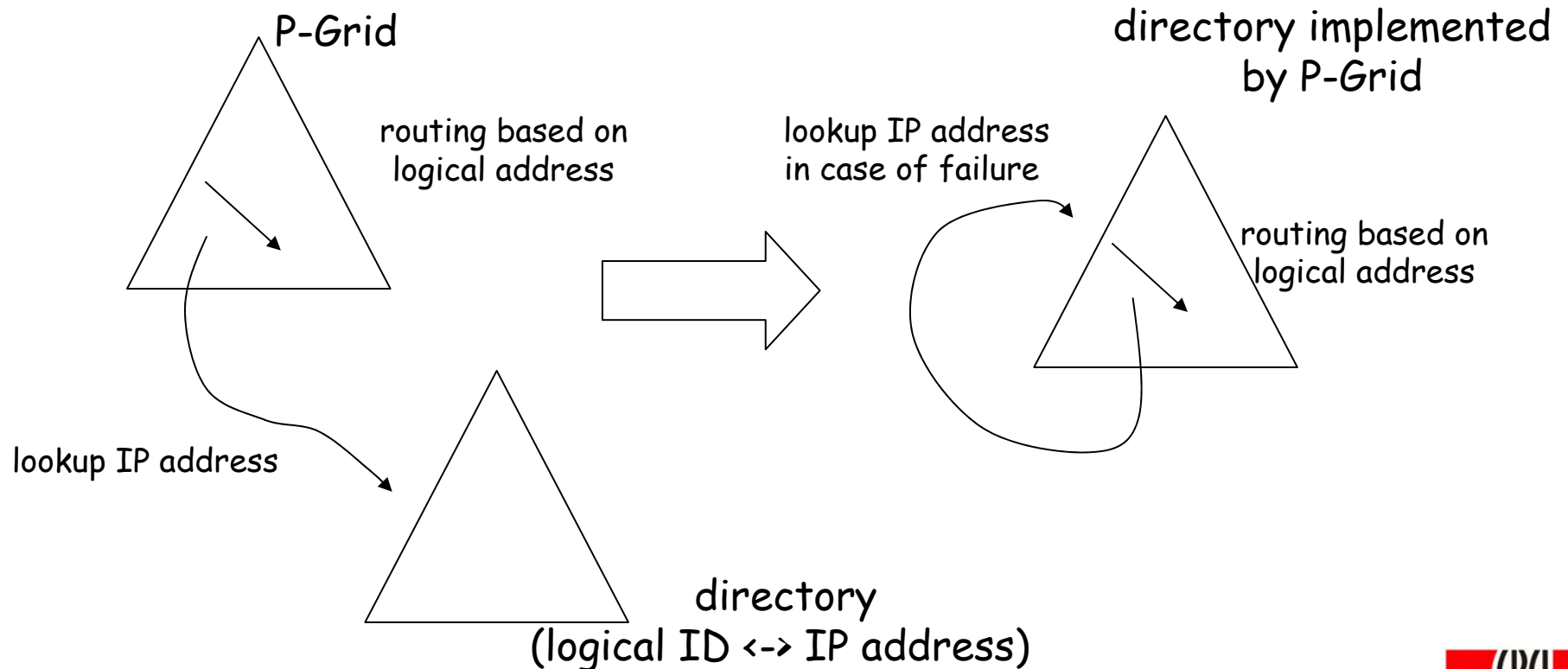
Scheme	Total messages per online peer	Push Rounds
Flooding + Partial Lists	35.22	5
Haas et al $G(0.8,2)$ [Infocomm 2002]	28.49	5
Our Scheme, $P_F(t)=0.8^t$	16.35	6

3. P-Grid: scalable, self-organizing distributed data access

4. Updates in a P2P environment

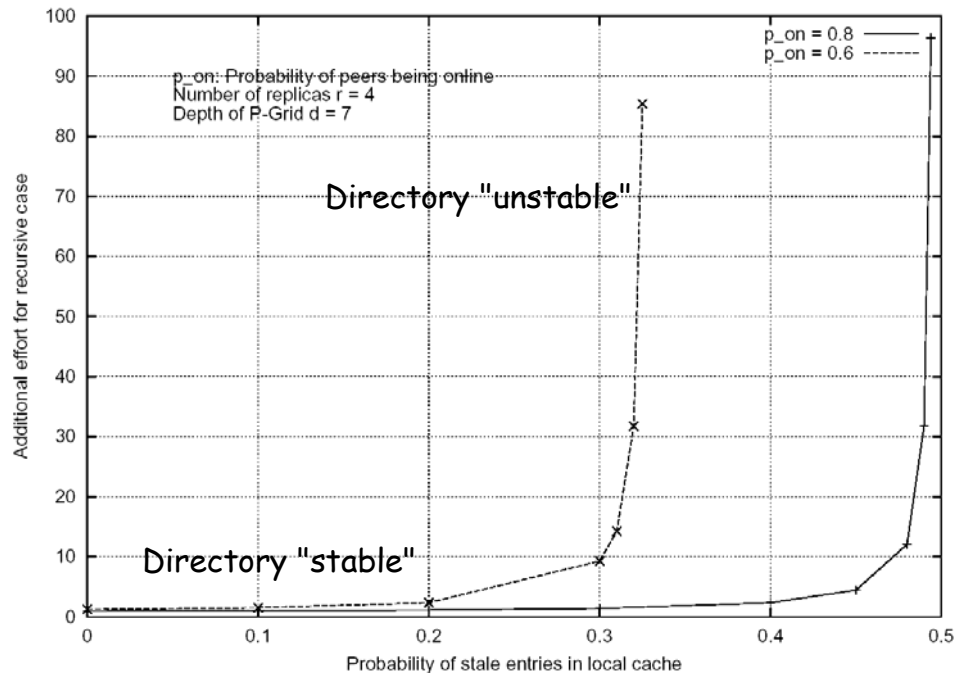
5. Application: Self-healing Directory

- Any distributed access structure (such as P-Grid) requires dynamic mapping of a logical ID (associated data key) to physical ID (IP address)
 - This mapping cannot be static in the presence of dynamic IP addresses
 - A very important problem for the implementation of any P2P system



Approach and Result

1. Store the mapping in a P-Grid access structure
2. Incoming peers register and the mapping is updated
3. If they reuse a logical ID they proof their identity using a secret key
4. P-Grid nodes cache mapping in their routing tables and use it for routing
5. In case of invalid cache entries recursively searches on the directory are initiated to repair the invalid cache



Results from analytical model

for details:
Manfred Hauswirth, Anwitaman Datta, Karl Aberer:
"Handling Identity in Peer-to-Peer Systems",
EPFL Technical Report IC/2002/67

6. Conclusion

Self-organizing Distributed Data Access

1. Construction of distributed data access structure as self-organizing process
2. Scalable, randomized, distributed algorithm for search, update exploiting 1.

Self-organizing Applications (Multi-Level Self-organization)

- Bootstrapping identity
- Reputation and trust, relevance and ranking

- Data management issues (in particular with semantically meaningful data)
 - skewed data distribution
 - updates
 - complex search predicates
- P-Grid implementation under way
 - Java, nice GUI, Gnutella/JXTA compatible

References

P-Grid: A Self-Organizing Access Structure for P2P Information Systems

Karl Aberer, *Sixth International Conference on Cooperative Information Systems (CoopIS 2001), Trento, Italy, Lecture Notes in Computer Science 2172, Springer Verlag, Heidelberg, 2001.*

Managing Trust in a Peer-2-Peer Information System

Karl Aberer, Zoran Despotovic, *10th International Conference on Information and Knowledge Management, (2001 ACM CIKM), ACM Press, New York, 2001.*

Improving Data Access in P2P Systems

Karl Aberer, Manfred Hauswirth, Magdalena Puceva, Roman Schmidt, *IEEE Internet Computing, 6(1), January/February 2002.*

Scalable Data Access in P2P Systems Using Unbalanced Search Trees

Karl Aberer, *Workshop on Distributed Data and Structures (WDAS-2002), Paris, France, 2002.*

A Decentralized Architecture for Adaptive Media Dissemination

Philippe Cudré-Mauroux, Karl Aberer, *IEEE International Conference on Multimedia and Expo (ICME2002), Lausanne, Switzerland, 2002.*

Updates in Highly Unreliable, Replicated Peer-to-Peer Systems

Anwitaman Datta, Manfred Hauswirth, Karl Aberer, *to be published in the proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS 2003), May 19-22, 2003, Providence, Rhode Island, USA.*

Gridella: an open and efficient Gnutella-compatible Peer-to-Peer System based on the P-Grid approach

Roman Schmidt, *EPFL Technical Report IC/2002/71, Master's Thesis at the Technical University of Vienna.*

Advanced Peer-to-Peer Networking: The P-Grid System and its Applications

Karl Aberer, Philippe Cudré-Mauroux, Anwitaman Datta, Zoran Despotovic, Manfred Hauswirth, Magdalena Puceva, Roman Schmidt, Jie Wu, *to be published in PIK journal.*

Handling Identity in Peer-to-Peer Systems

Manfred Hauswirth, Anwitaman Datta, Karl Aberer, *6th International Workshop on Mobility in Databases and Distributed Systems, MDDS'2003, in conjunction with the 14th International Conference on Database and Expert Systems Applications DEXA'2003, September 1-5, 2003, Prague, Czech Republic*

Efficient Search in Unbalanced, Randomized Peer-To-Peer Search Trees

Karl Aberer, *EPFL Technical Report IC/2002/79.*

Self-organized construction of distributed access structures: A comparative evaluation of P-Grid and FreeNet

Karl Aberer, Manfred Hauswirth, Magdalena Puceva, *EPFL Technical Report IC/2002/74.*

Beyond "Web of Trust": Enabling P2P E-commerce

Anwitaman Datta, Manfred Hauswirth, Karl Aberer, *IEEE Conference on E-Commerce, CEC'03, June 24-27, 2003, Newport Beach, California, USA.*