
# Information Sharing in Peer-to-Peer Systems

Nuno Lopes

Departamento de Informática

Novembro, 2006


# Contents

# P2P System Characterization

- Large number of hosts.
- Completely decentralized.
- Highly variable host uptime.

# Searching Information on P2P systems

- Unstructured overlays rely on flood-based queries (Gnutella)
- Structured overlays rely on distributed hash tables (Chord, Pastry,...)

# Search on Distributed Hash Table systems

- Efficient object location given a key.
- No explicit search functionality, only exact key matching.
- Inverted index model used to implement searching: $word \hookrightarrow document\_reference_{set}$.
- Unbalanced load caused by Zipf distribution of data.


# Contents

# Distributed Balanced Tree Algorithm

- Operations have logarithmic cost, following tree depth.
- Block independence adapts to dictionary-like interface storage (was originally designed for secondary storage).
- Decentralized unique block key generation scheme.
- Related work presents other tree-based structures.

## Generic P2P Index

- Tree structure can be used as a Set implementation, to build a single large-scale textual inverted index for example.
- It can also be used for any distributed application that would benefit from a large-scale $B^+$-Tree implementation.

# Customized Discrete Event Simulator

- A Python-based simulator was developed.
- It follows the “Scalable Simulation Framework” model for network oriented simulations.
- It provides with a simpler API and library interface than it’s Java counterpart.
- To be made available to the community in the future.

# Contents

# Open Issues

- Evaluating performance of the algorithm for range-query implementation.
- Semantic-based data reconciliation algorithm is being developed to handle data consistency issues.

## Final Remarks

- This thesis presents a solution to store and search for unbalanced data over any DHT system without losing scalability.
- Furthermore, by using a well known data algorithm, the $B^+$ Tree, a generic index functionality is available for developing new large-scale P2P applications.