Efficient and decentralized discovery of approximate global state

S. Keshav

June 2005

Need for global state

- Sensor field: compute min, max, ave
- P2P network: find popular items
- Stream database: find top-K items
- Internet routing: find best interface for destination
- Today's talks
  - BGP policies
  - Channel and power assignment
  - DOMINO data sharing

System assumptions

- Large number of nodes
  - nodes join and leave
  - links may fail
- Computation may be massively distributed
- Values at each node change over time

Model

- N nodes
- State at node i is s(i,t)
- S = {s(1,t), s(2, t) ... s(N,t)}
- Compute f (S,t)
  - [Bawa et al 2004]

f may be incomputable

- f is well defined
  - but may be uncomputable
- Consider a node that sends data, then dies
- And the data is lost!

However...

- In practice, f computed over large enough subset of N should be sufficient
- Thus, approximate computation of global state

Some more structure...

- Taxonomy
- Metrics
- Solution approaches

Taxonomy: function

- Function being computed
  - Extremal
  - Histogram
  - Measure of central tendency
  - Routing table
  - Policy
  - Optimal channel allocation

Taxonomy: topology

- Network topology
  - Clique
  - Random (k)
  - Tree (k)
  - Hypercube
  - PLRG/Hierarchical PLRG
  - Real internet - Rocketfuel

Taxonomy: change model

- State change model
  - Change in node state
  - Nodes join or leave
  - Links go up and down

Metrics

- Accuracy
- Cost
- Speed
- Robustness
- Scalability

Solution approaches

- Centralization
- Tree-based
- Random walk
- Randomized gossip

Centralization and tree-based approaches

- Fast
- Accurate
- Low cost
- But not scalable or robust...

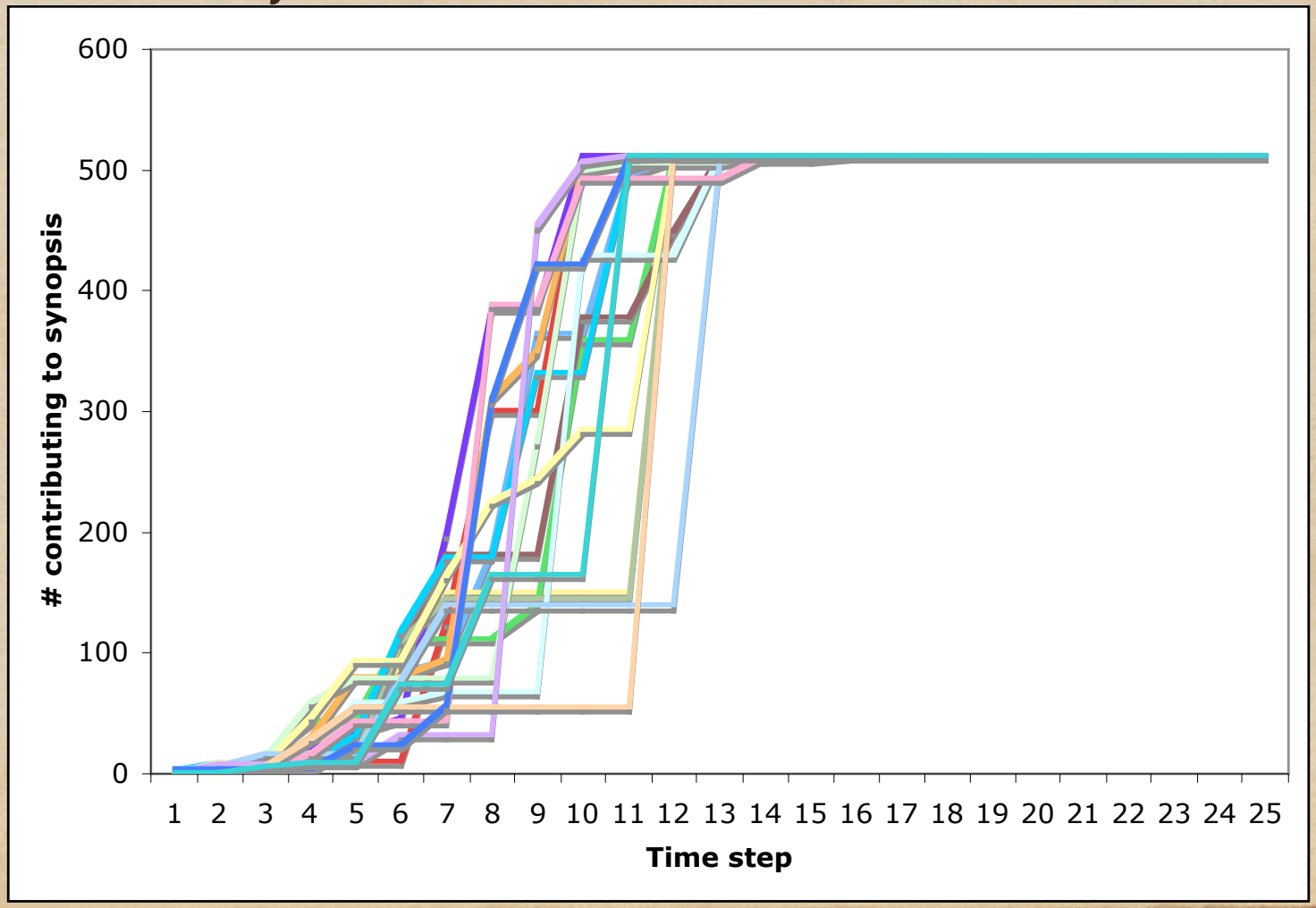
Randomized approaches

- Fast: O(log N + 1/error_bound) time
- Low cost
- Robust - no need for error recovery
- Accuracy depends on the scheme, but usually probabilistic
- Scalable
- But -- need to avoid duplication

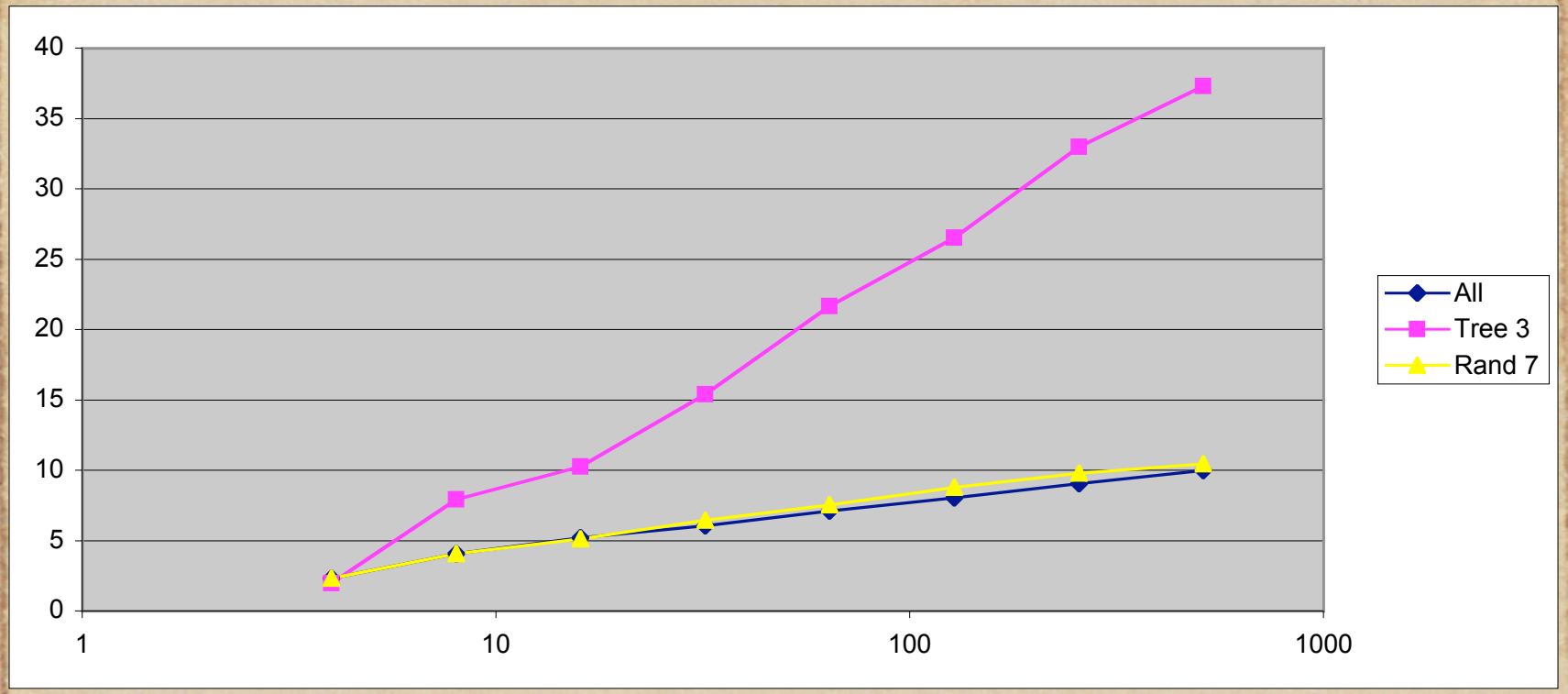
Avoiding duplication

- Duplicate insensitive statistics (ODI)
  - Convert count to extremal value [Nath]
- Mass conservation
  - 'Push-synopsis' [KDG 03]
- Tag statistics with ID of node adding information
  - Need solve scaling problem

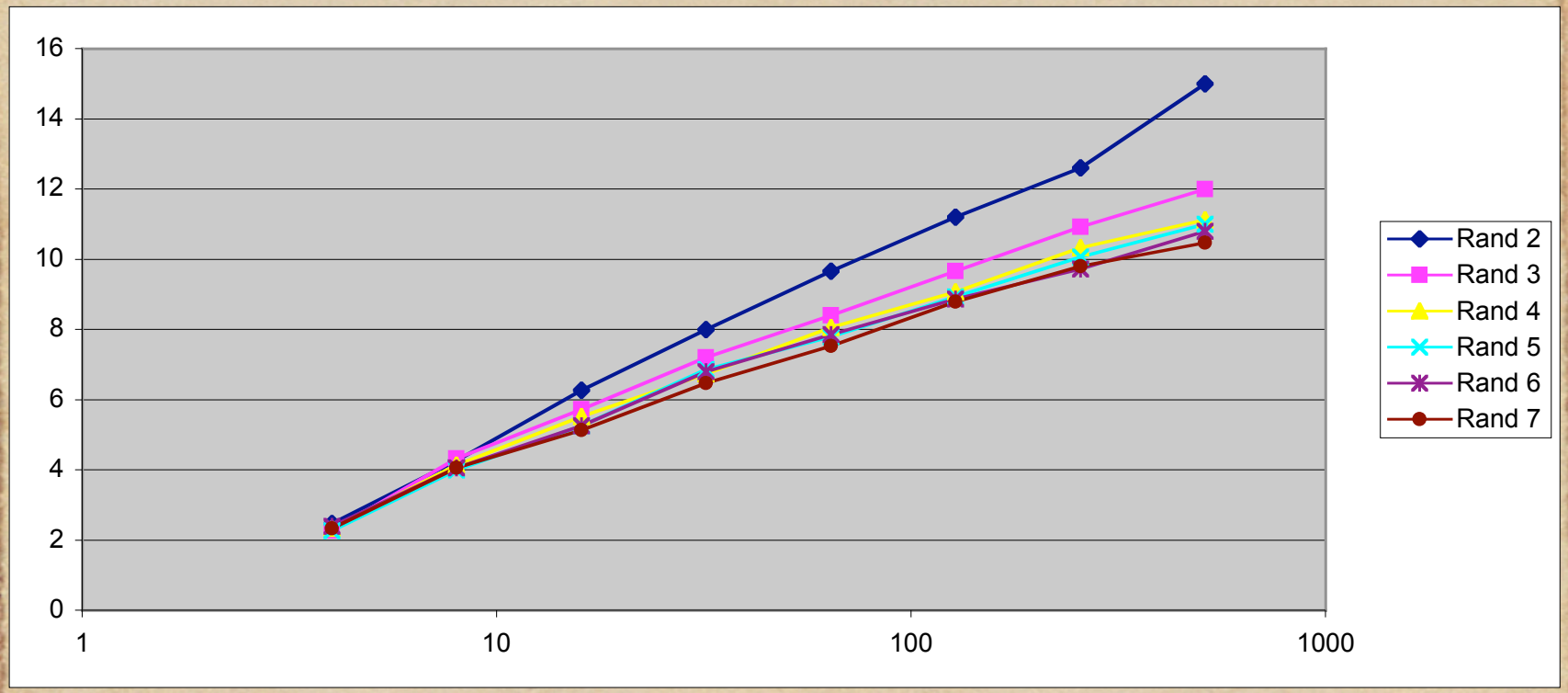
Sample of convergence



Convergence with other topologies



Effect of # neighbours



Open problems

- Which approach is 'best'?
- How to model real problems (routing?)
- Practical considerations
  - detecting termination
  - fault tolerance
  - sensitivity to topology
  - removing staleness
  - security

Challenge

- If we can solve these problems, then it opens up a new approach to distributed self-organization
- At the intersection of distributed systems, networking, and databases!