Data Storage Solutions for Decentralized Online Social Networks

— Anwitaman Datta

S* Aspects of Networked & Distributed Systems (SANDS)
School of Computer Engineering
NTU Singapore

iSocial Summer School, KTH Stockholm

Research @ SANDS

Applications

recommendation and decision support systems

decentralized online social networking and collaboration

(Distributed) Systems

privacy aware/preserved data aggregation, storage, sharing & analytics/data-mining

data/computation at 3rd party/outsourced

distributed key-value stores

data-center design

P2P/F2F storage systems

networked distributed storage & data management systems

Foundational

social network analysis

trust models

secure/privacy preserved computation primitives

codes for storage

DOSNish research at SANDS

DOSNish research at SANDS

- Selective information dissemination using social links
  - GoDisco

DOSNish research at SANDS

- Selective information dissemination using social links
  - GoDisco
- Security issues
  - Access control, Private Information Retrieval, …

DOSNish research at SANDS

- Selective information dissemination using social links
  - GoDisco
- Security issues
  - Access control, Private Information Retrieval, …
- DOSN architectures
  - PeerSoN, SuperNova, PriSM, …

DOSNish research at SANDS

- Selective information dissemination using social links
  - GoDisco
- Security issues
  - Access control, Private Information Retrieval, …
- DOSN architectures
  - PeerSoN, SuperNova, PriSM, …
- P2P storage

DOSNish research at SANDS

- Selective information dissemination using social links
  - GoDisco
- Security issues
  - Access control, Private Information Retrieval, …
- DOSN architectures
  - PeerSoN, SuperNova, PriSM, …
- P2P storage

http://sands.sce.ntu.edu.sg/dOSN

P2P Storage

- Not the same as a file-sharing system
  - Peer-to-Peer (P2P) storage systems leverage the combined storage capacity of a network of storage devices (peers) contributed typically by autonomous end-users as a common pool of storage space to store content reliably.

P2P Storage

P2P Storage

- Design space

P2P Storage

- Design space
  - Reliability: Availability & Durability (focus of this talk)

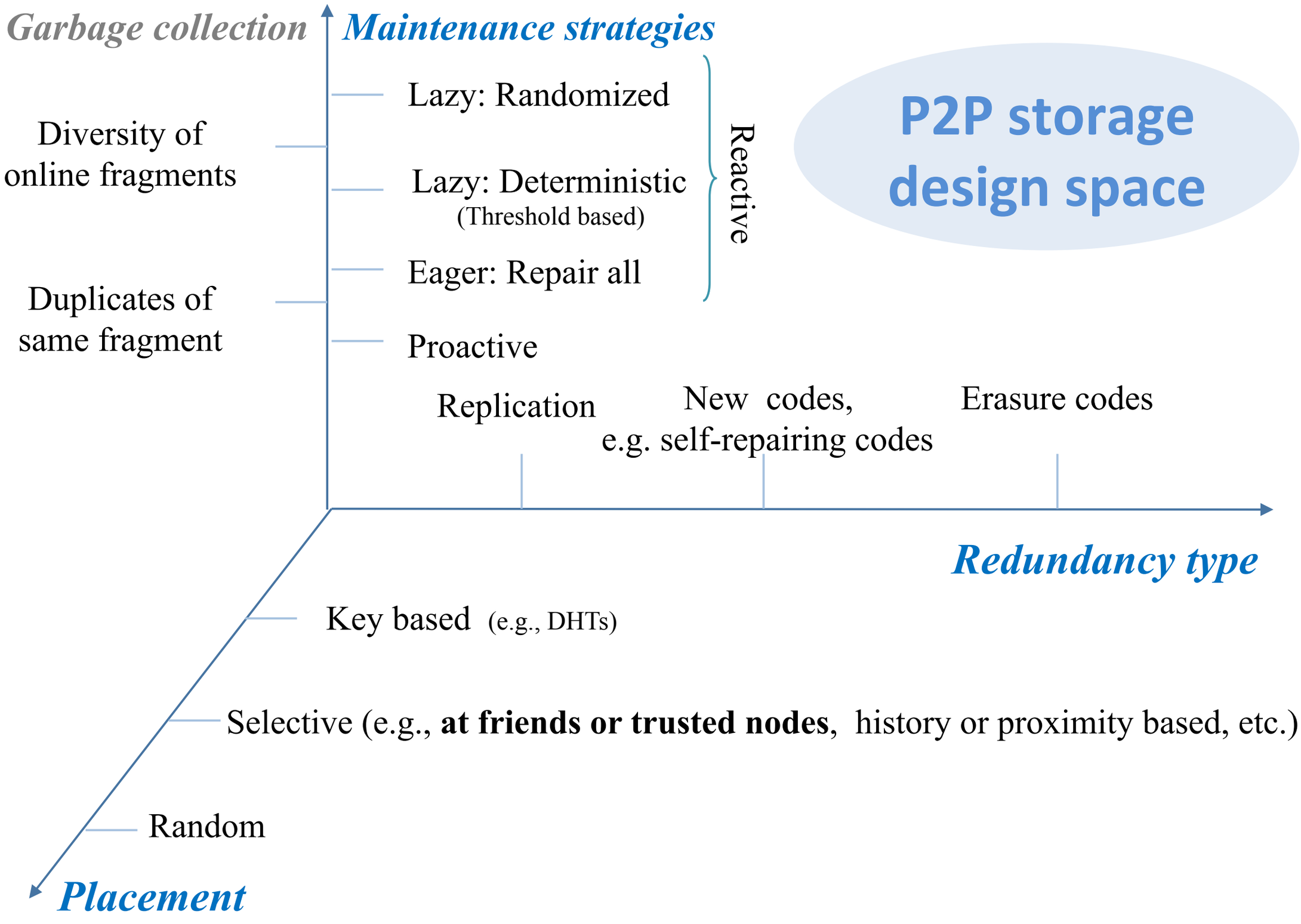
P2P Storage

- Design space
  - Reliability: Availability & Durability (focus of this talk)
  - Security & Privacy: Access control, integrity, free-riding, anonymity, privacy, …

P2P Storage

- Design space
  - Reliability: Availability & Durability (focus of this talk)
  - Security & Privacy: Access control, integrity, free-riding, anonymity, privacy, …
  - Sophisticated functionalities: Concurrency, Version Control, …

Realizing Reliability

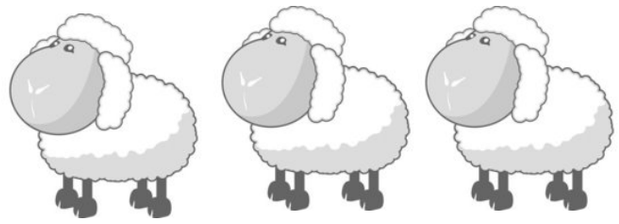


P2P storage design space

Redundancy Type

Redundancy Type

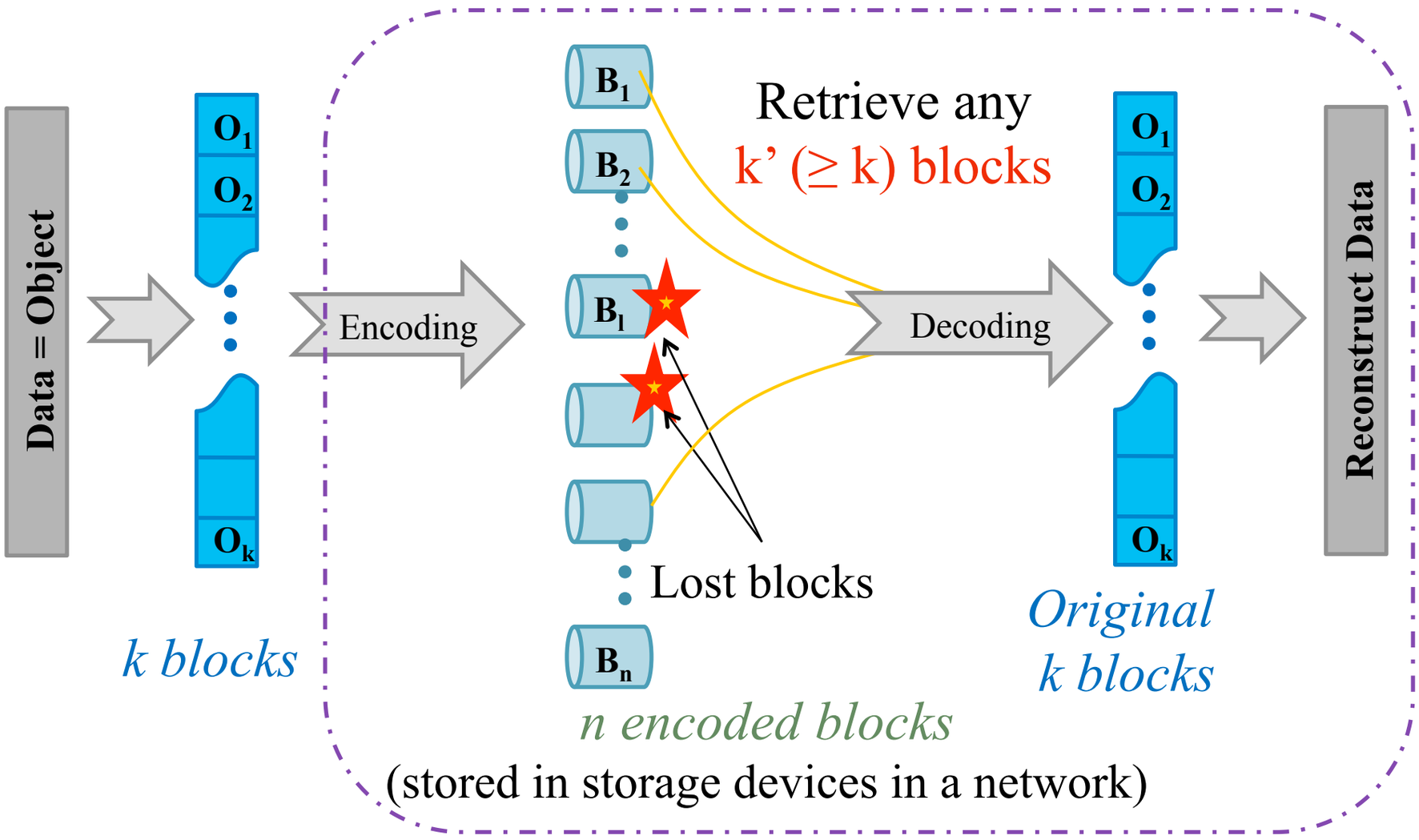
- Replication



Redundancy Type

- Replication

- Erasure codes



Redundancy placement

Redundancy placement

- A rather complicated problem
  - All peers are fully cooperative and altruistic, but autonomous
    - System capacity and resource allocation …
      - Heterogeneity, …
    - Coverage: history/prediction/…

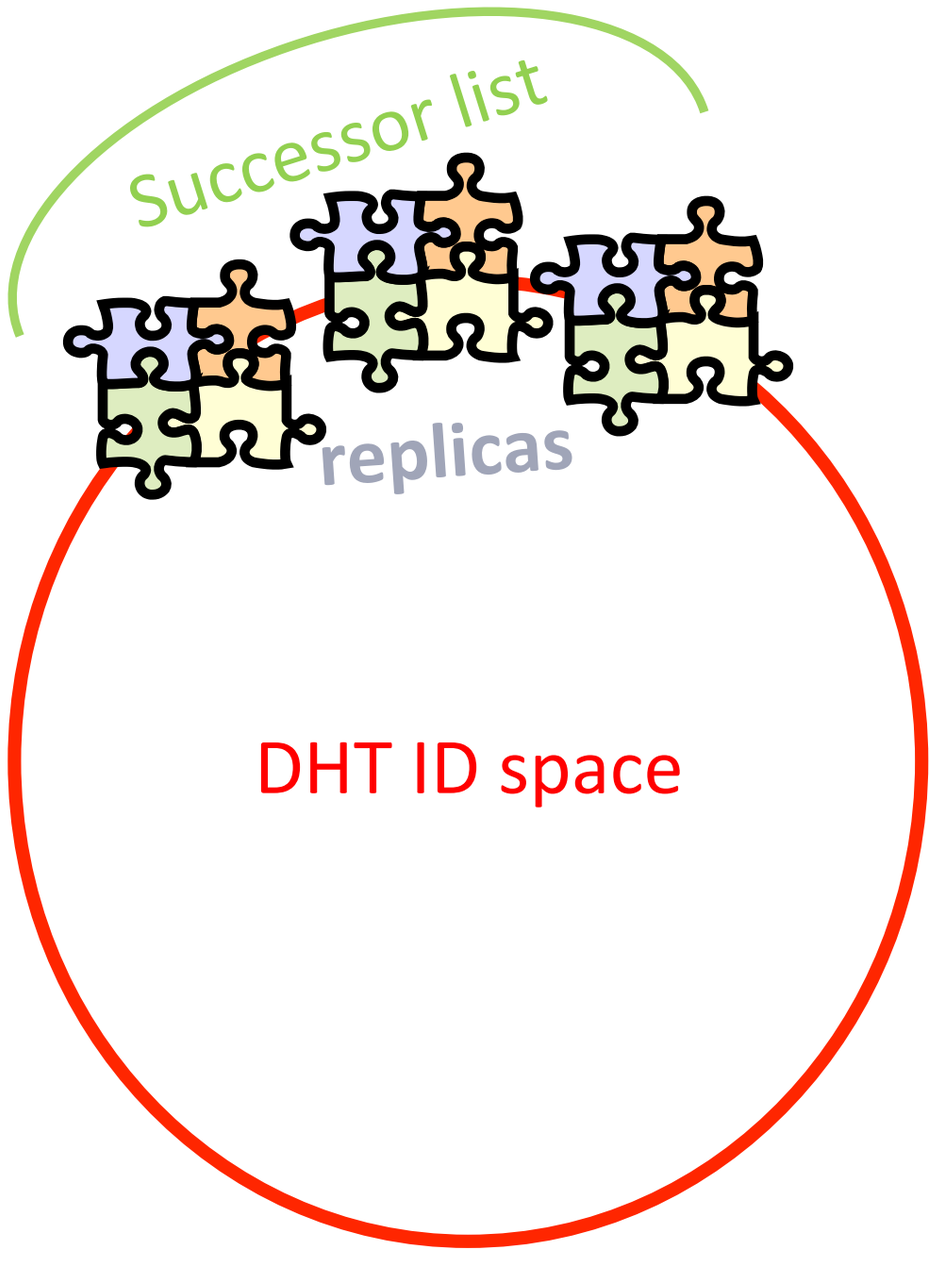
Redundancy placement

- A rather complicated problem
  - All peers are fully cooperative and altruistic, but autonomous
    - System capacity and resource allocation …
      - Heterogeneity, …
    - Coverage: history/prediction/…
  - Selfish/Byzantine peers: Incentives, trust, enforcement, …

Redundancy placement

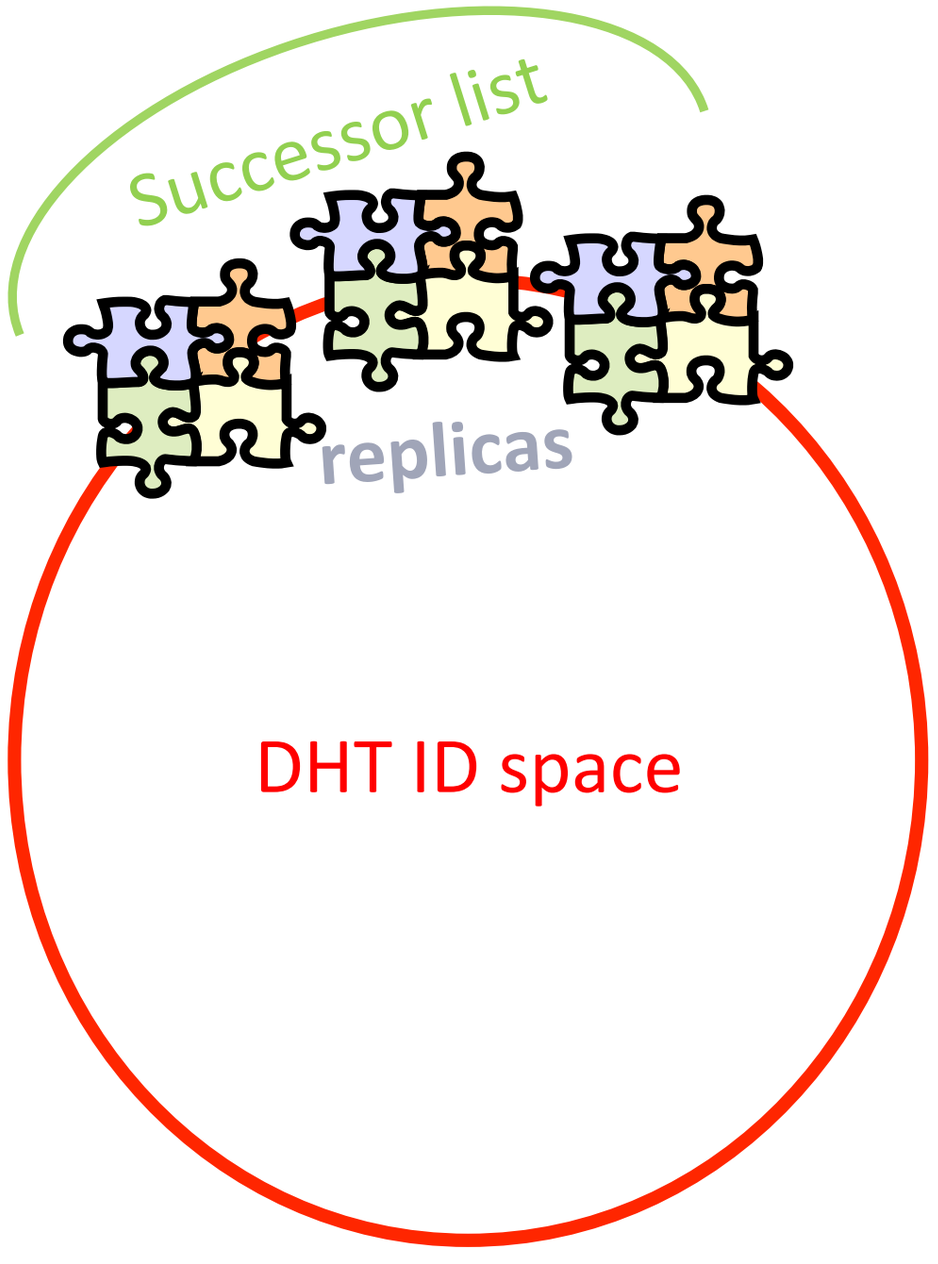
- A rather complicated problem
  - All peers are fully cooperative and altruistic, but autonomous
    - System capacity and resource allocation …
      - Heterogeneity, …
    - Coverage: history/prediction/…
  - Selfish/Byzantine peers: Incentives, trust, enforcement, …
  - Security & privacy implications of data placement …

Classical P2P storage systems



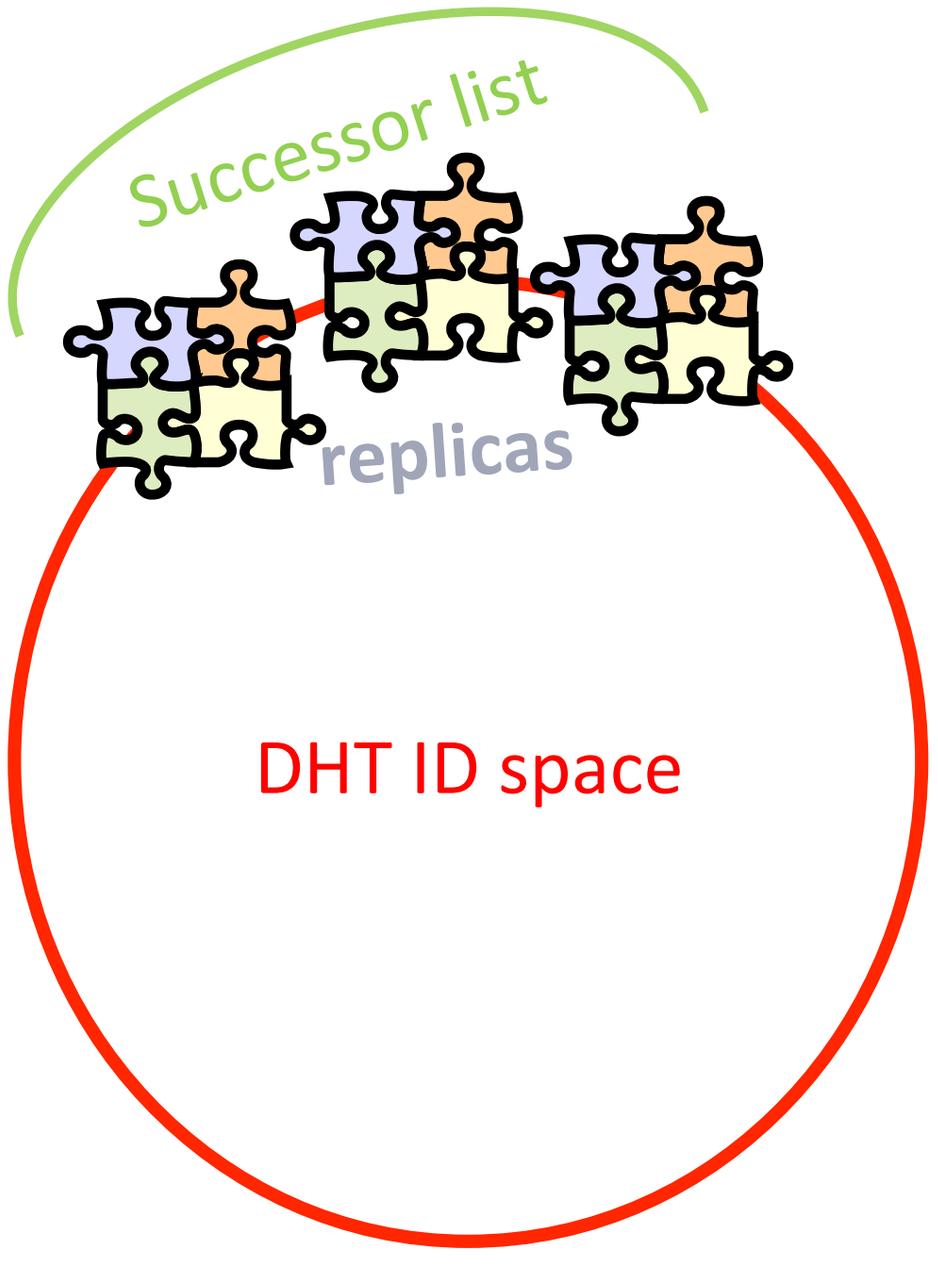
Classical P2P storage systems

- Distributed Hash Table (DHT) determines storage placement, e.g., CFS/ OpenDHT



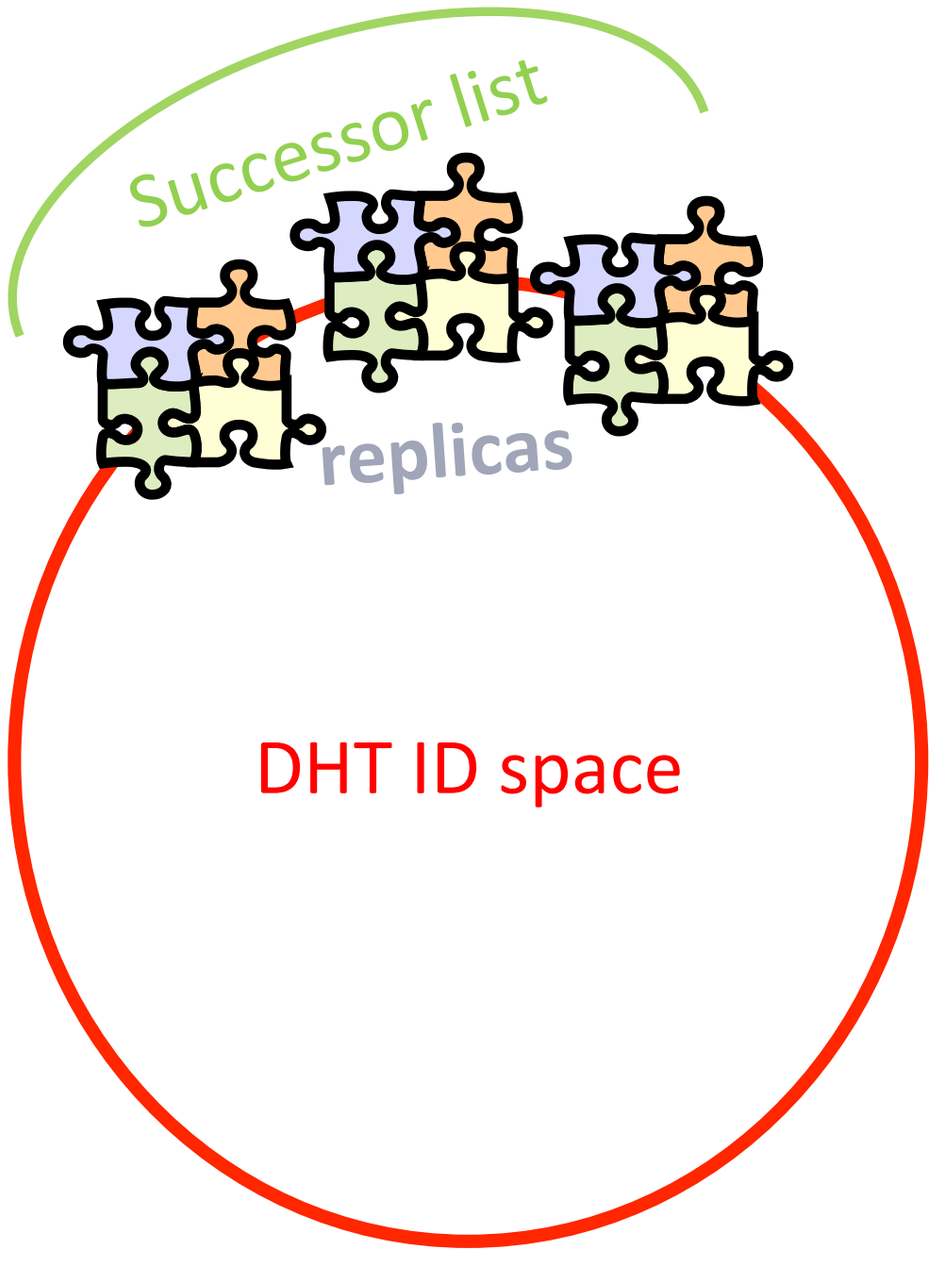
Classical P2P storage systems

- Distributed Hash Table (DHT) determines storage placement, e.g., CFS/ OpenDHT
  - Pros: Simple design, ease of locating data



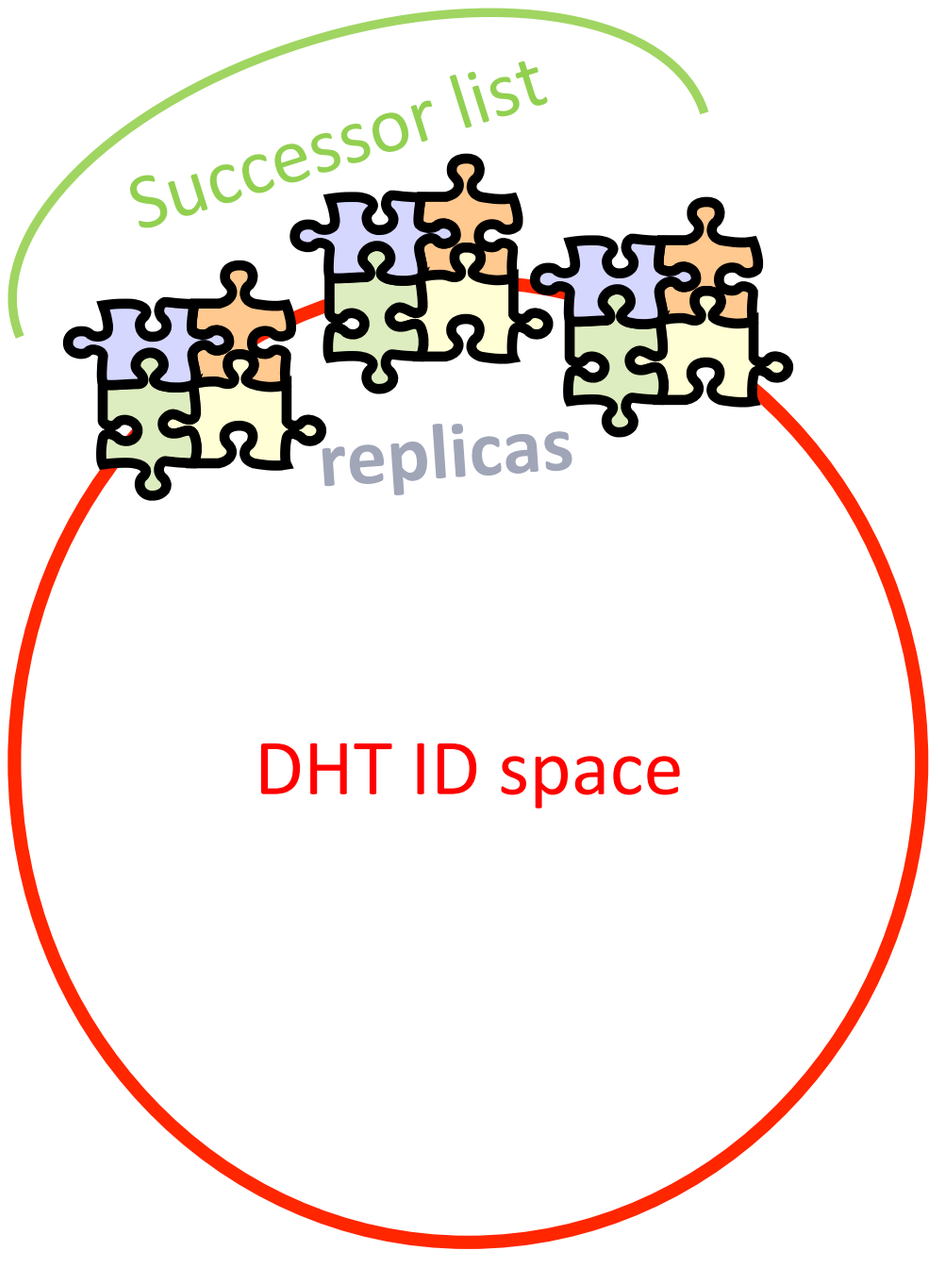
Classical P2P storage systems

- Distributed Hash Table (DHT) determines storage placement, e.g., CFS/ OpenDHT
  - Pros: Simple design, ease of locating data
  - Cons: mixes indexing with storage



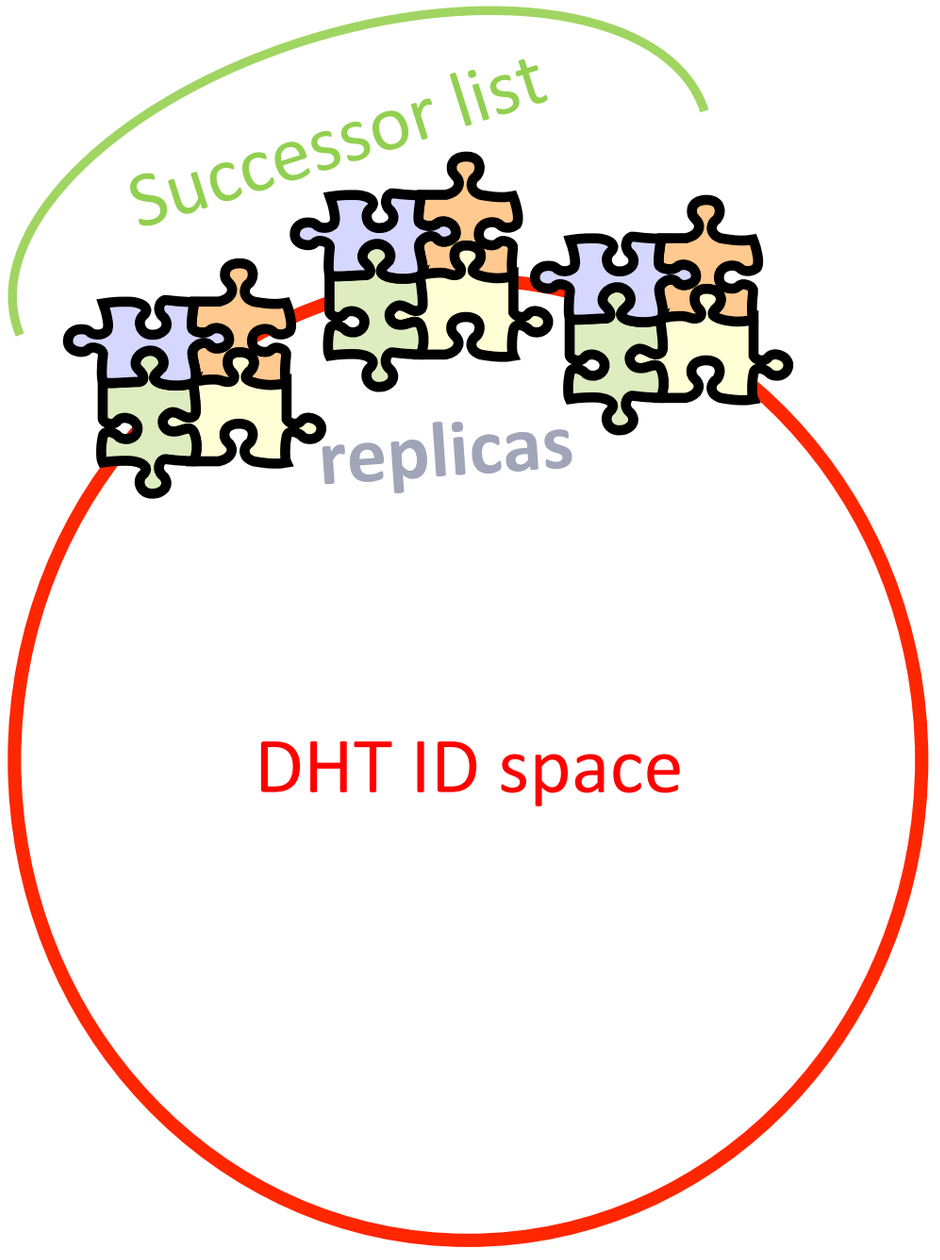
Classical P2P storage systems

- Distributed Hash Table (DHT) determines storage placement, e.g., CFS/ OpenDHT
  - Pros: Simple design, ease of locating data
  - Cons: mixes indexing with storage
    - high correlation of failures



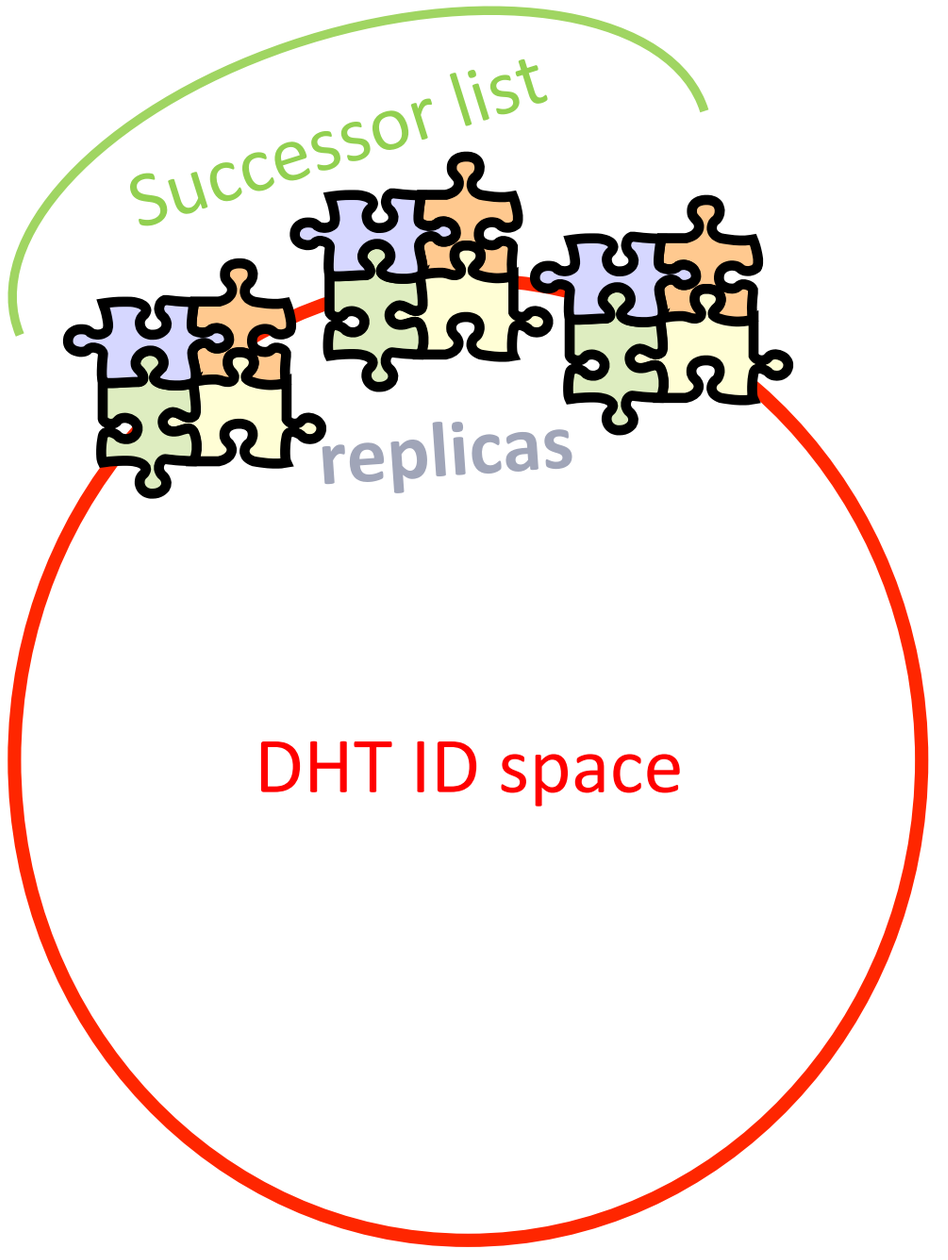
Classical P2P storage systems

- Distributed Hash Table (DHT) determines storage placement, e.g., CFS/ OpenDHT
  - Pros: Simple design, ease of locating data
  - Cons: mixes indexing with storage
    - high correlation of failures
    - cannot leverage other characteristics
      - e.g., locality, history, etc.

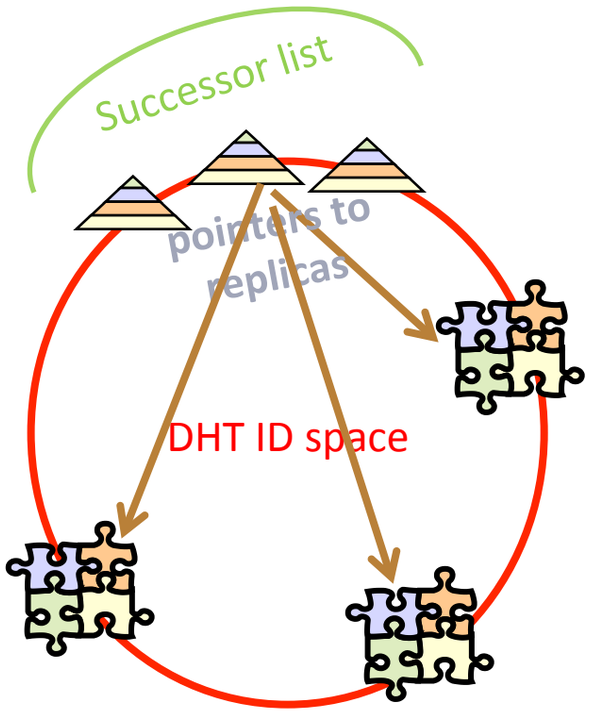


Classical P2P storage systems

- Distributed Hash Table (DHT) determines storage placement, e.g., CFS/ OpenDHT
  - Pros: Simple design, ease of locating data
  - Cons: mixes indexing with storage
    - high correlation of failures
    - cannot leverage other characteristics
      - e.g., locality, history, etc.
    - may lead to poor performance
      - access latency, repair cost, …

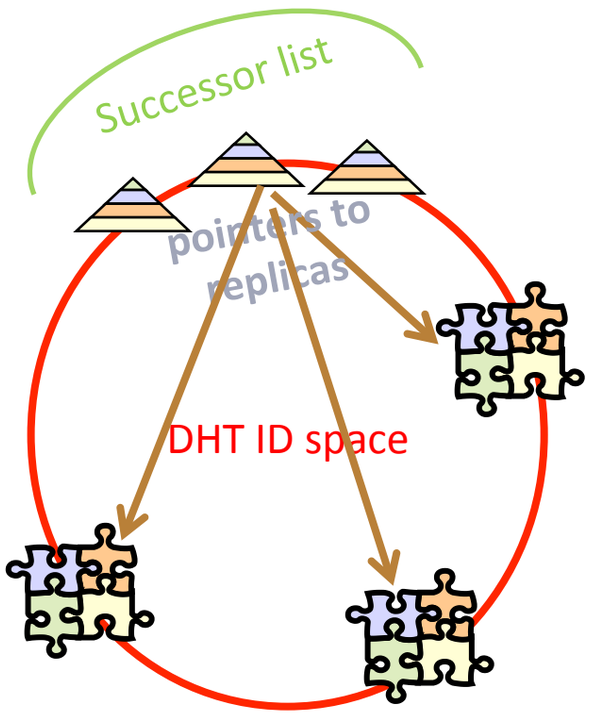


Classical P2P storage systems



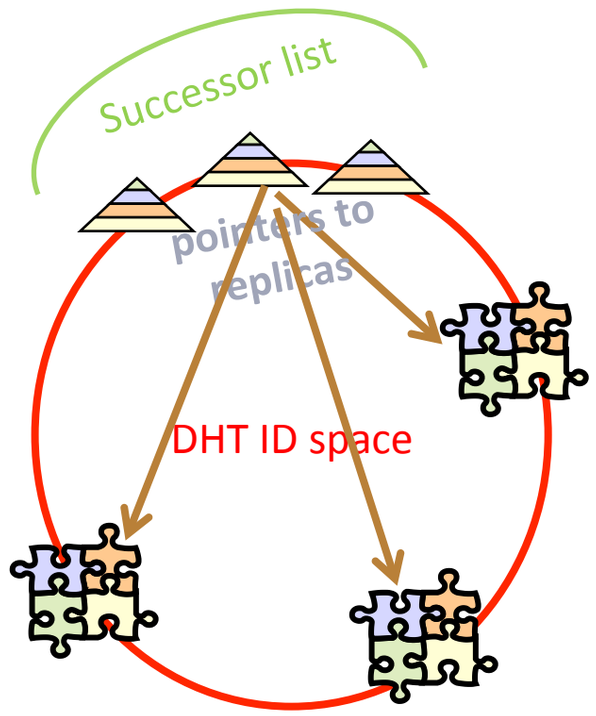
Classical P2P storage systems

- Distributed Hash Table (DHT) as a directory, e.g., TotalRecall



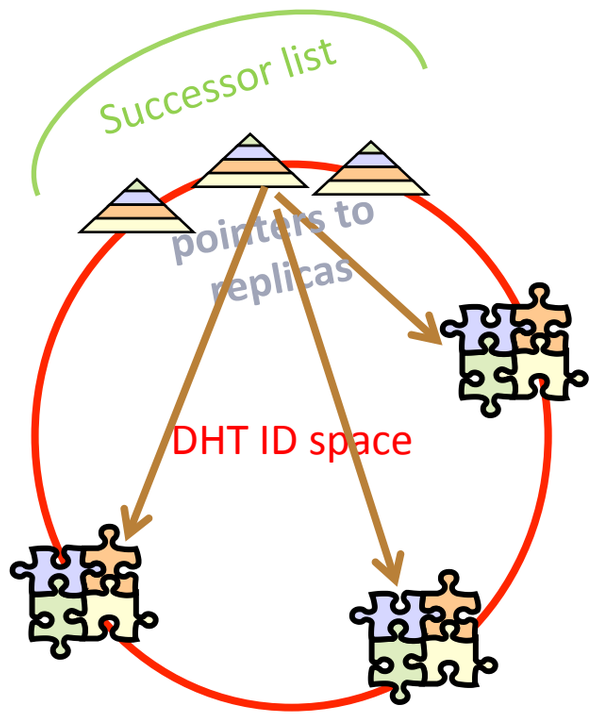
Classical P2P storage systems

- Distributed Hash Table (DHT) as a directory, e.g., TotalRecall
  - Pros: Flexible placement policy



Classical P2P storage systems

- Distributed Hash Table (DHT) as a directory, e.g., TotalRecall
  - Pros: Flexible placement policy
  - Cons of TotalRecall, which placed at random:
    - ???



Cloud assisted storage system

Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

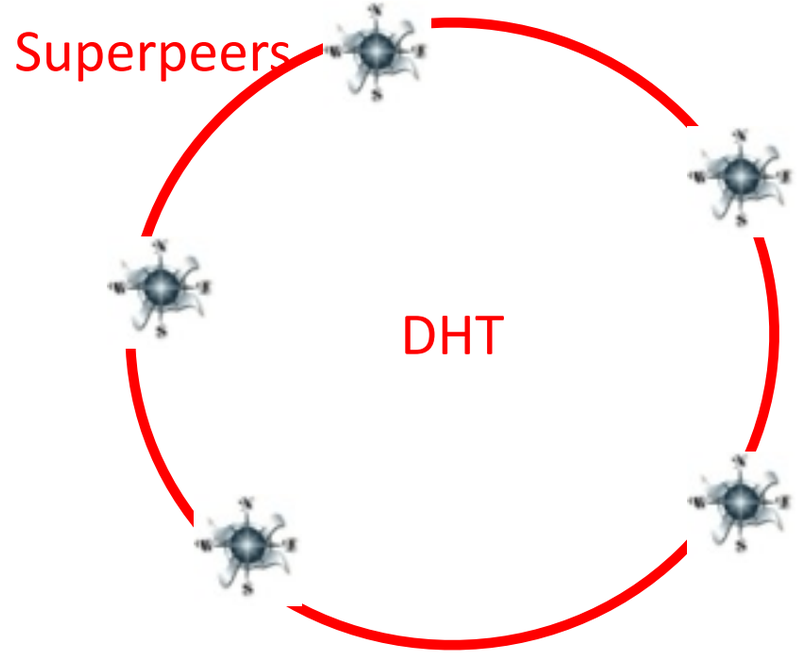
Cloud assisted storage system

- Hybrid architecture (used previously in Wuala)

Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

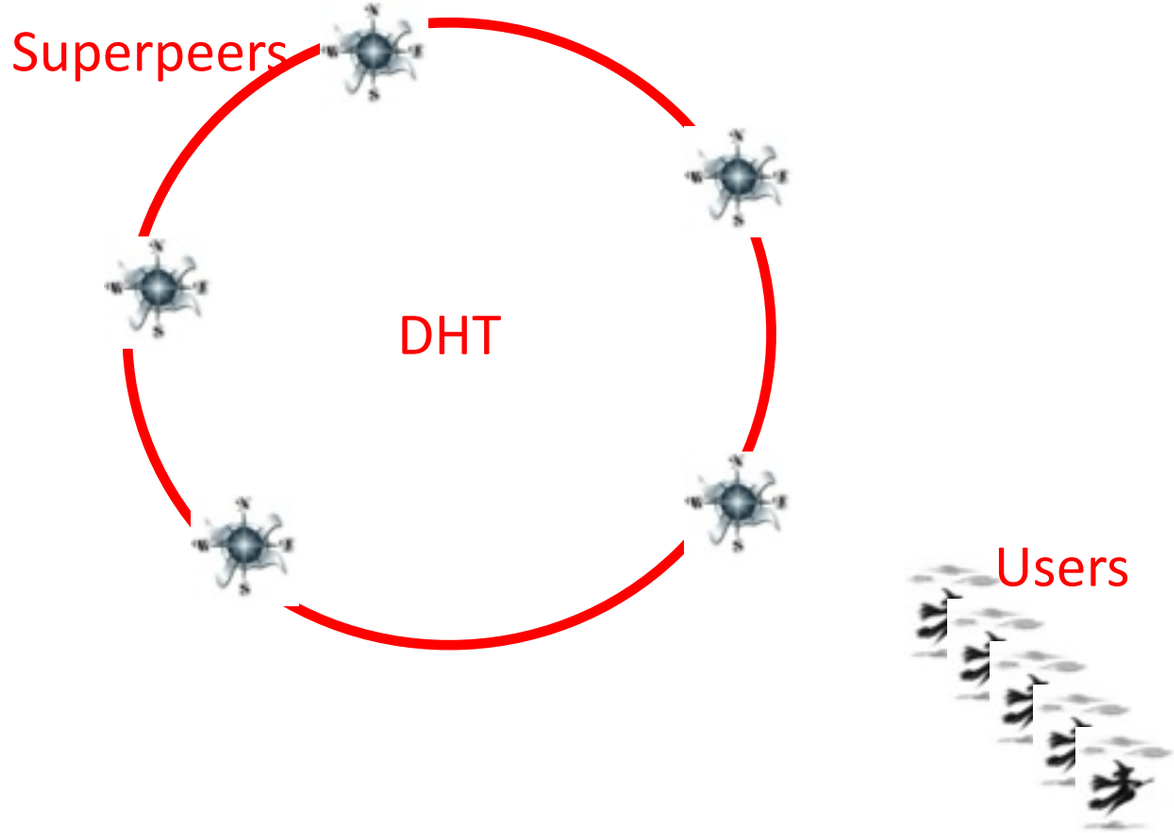
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

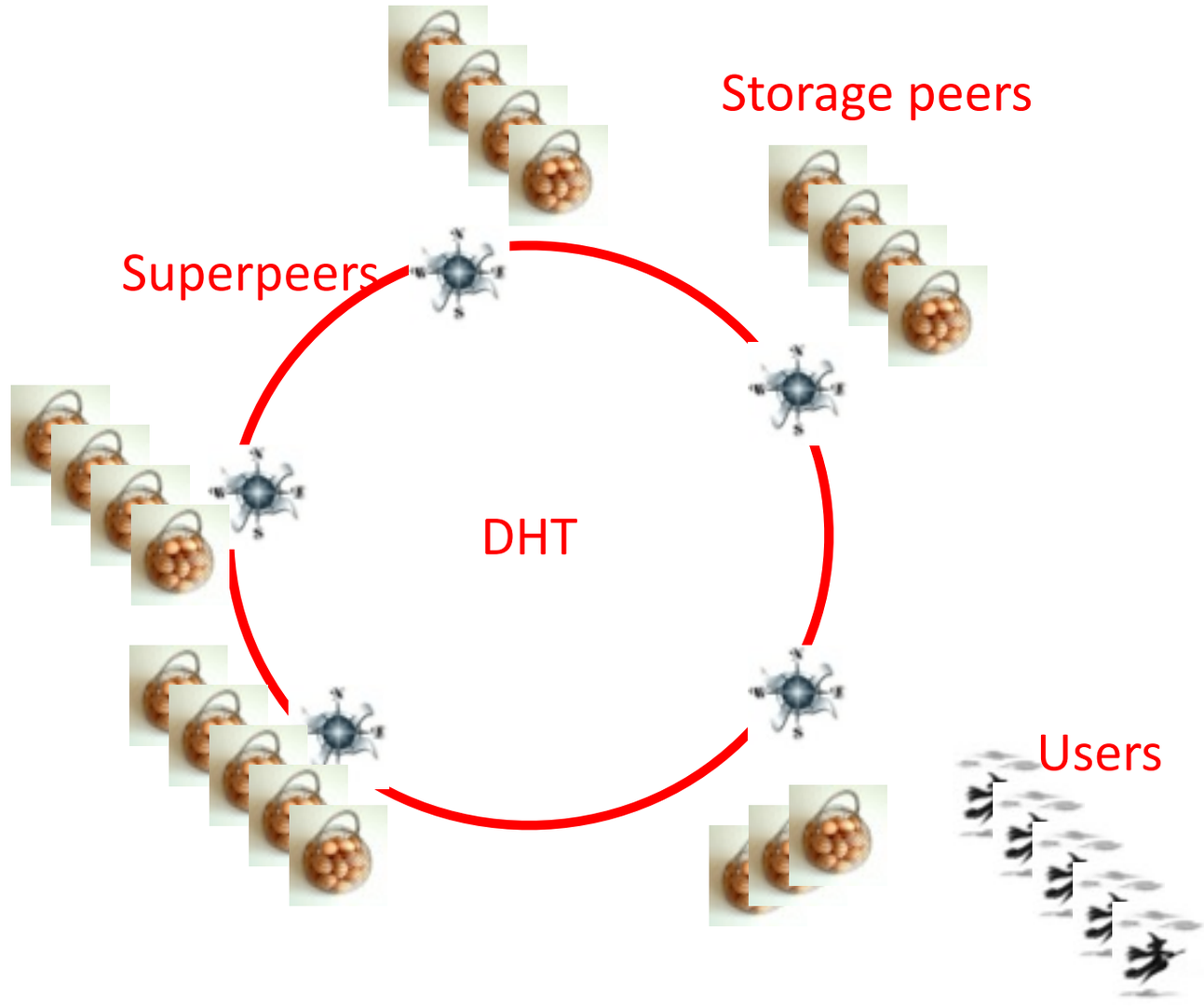
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

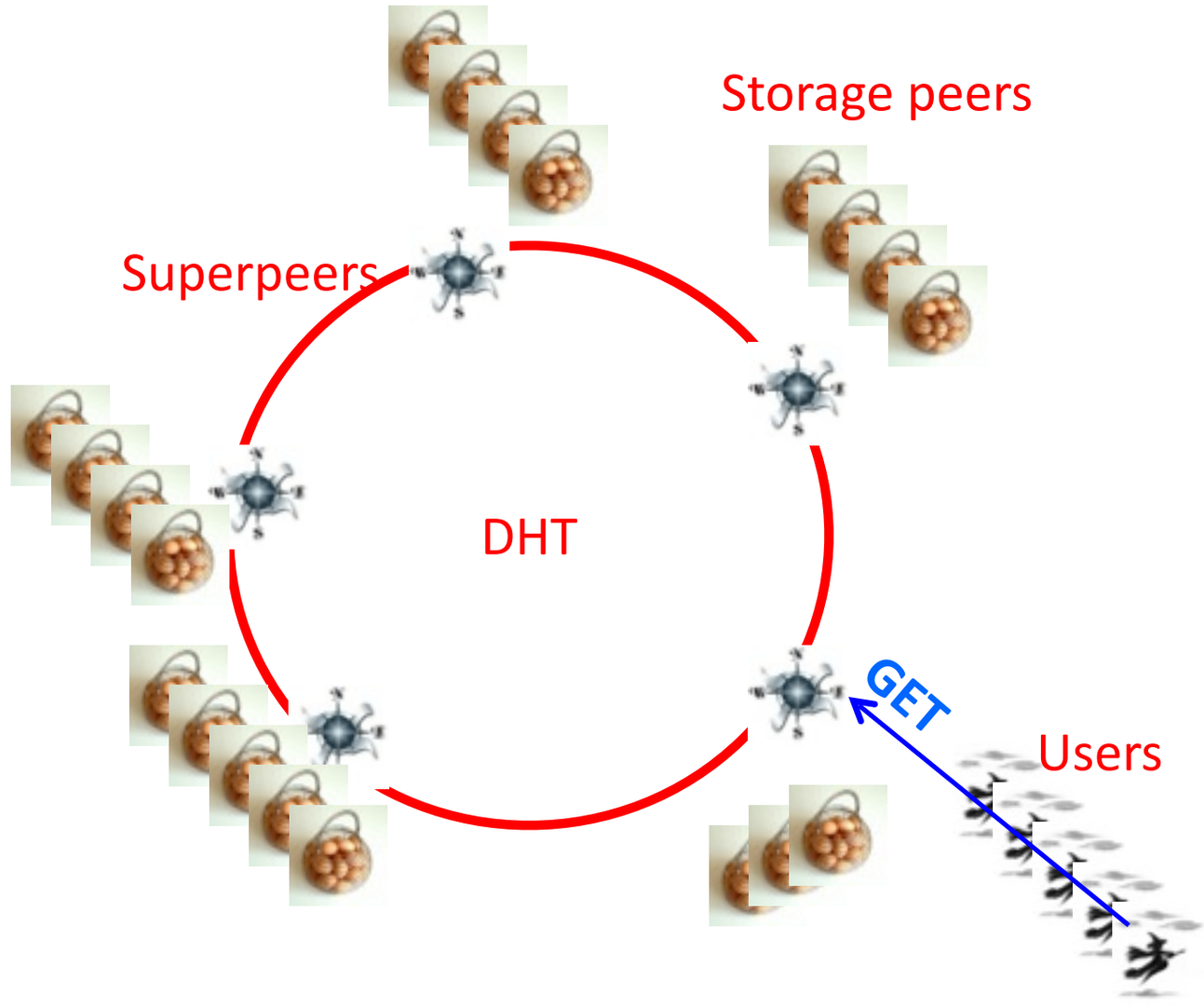
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

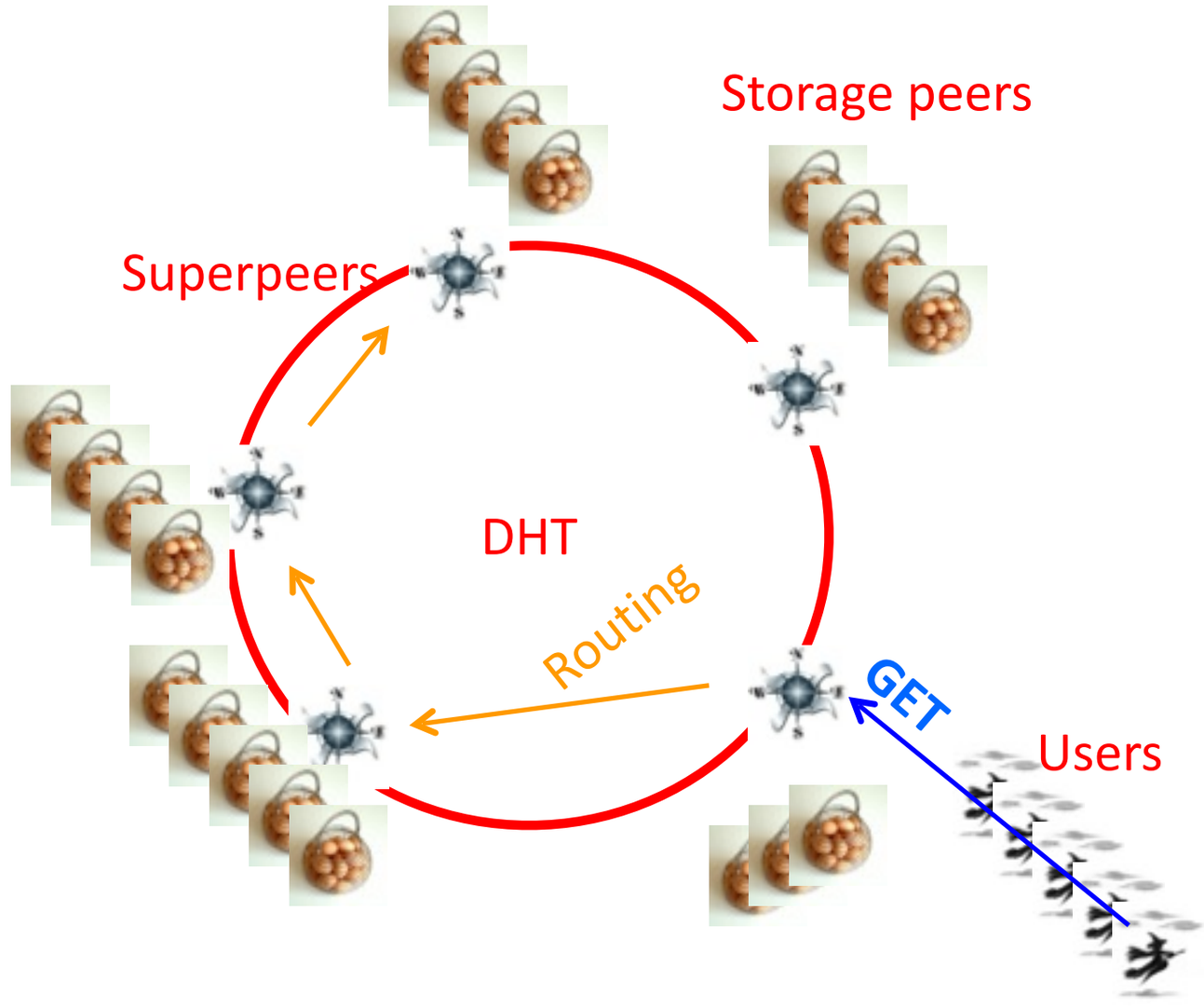
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

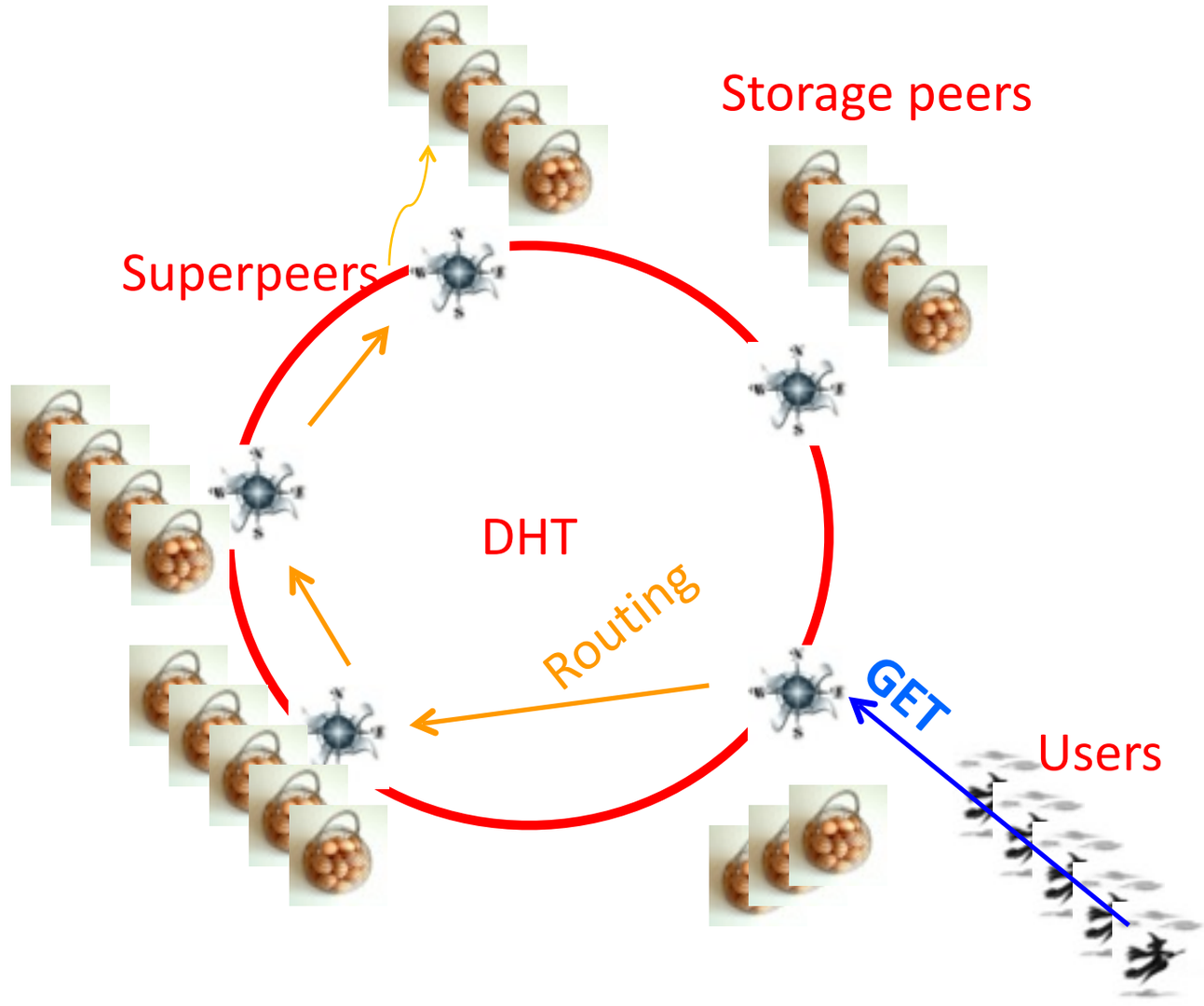
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

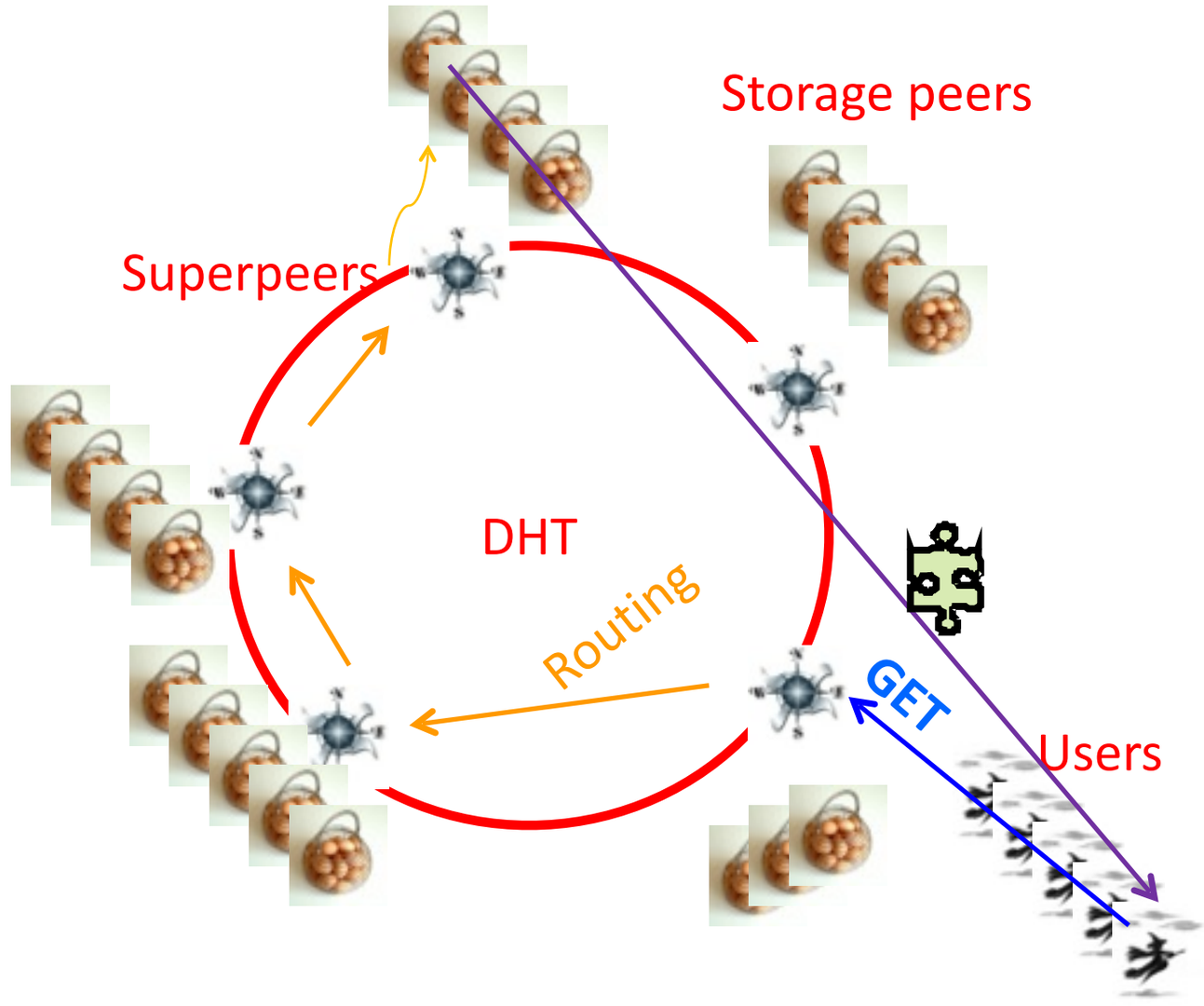
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

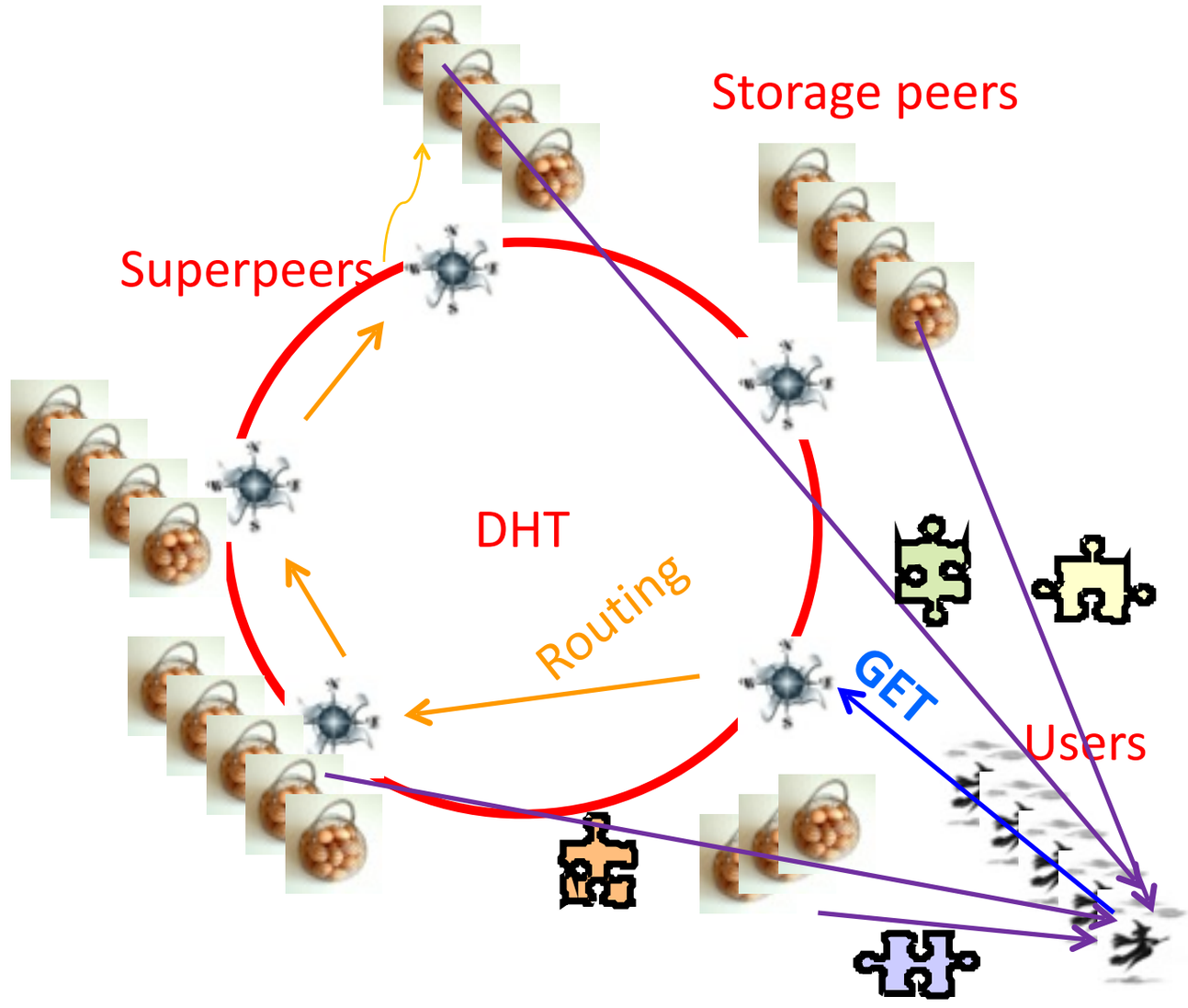
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

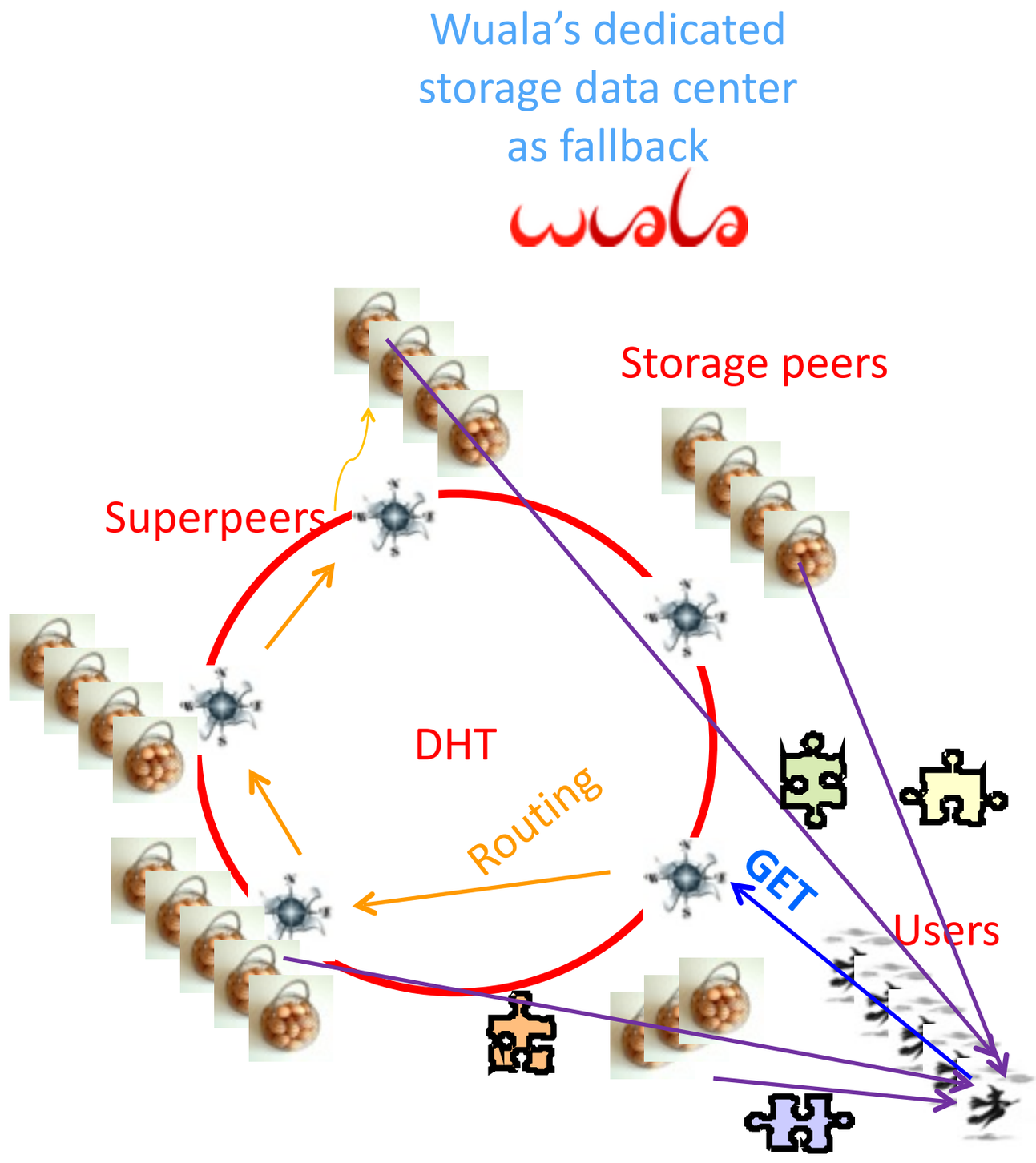
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

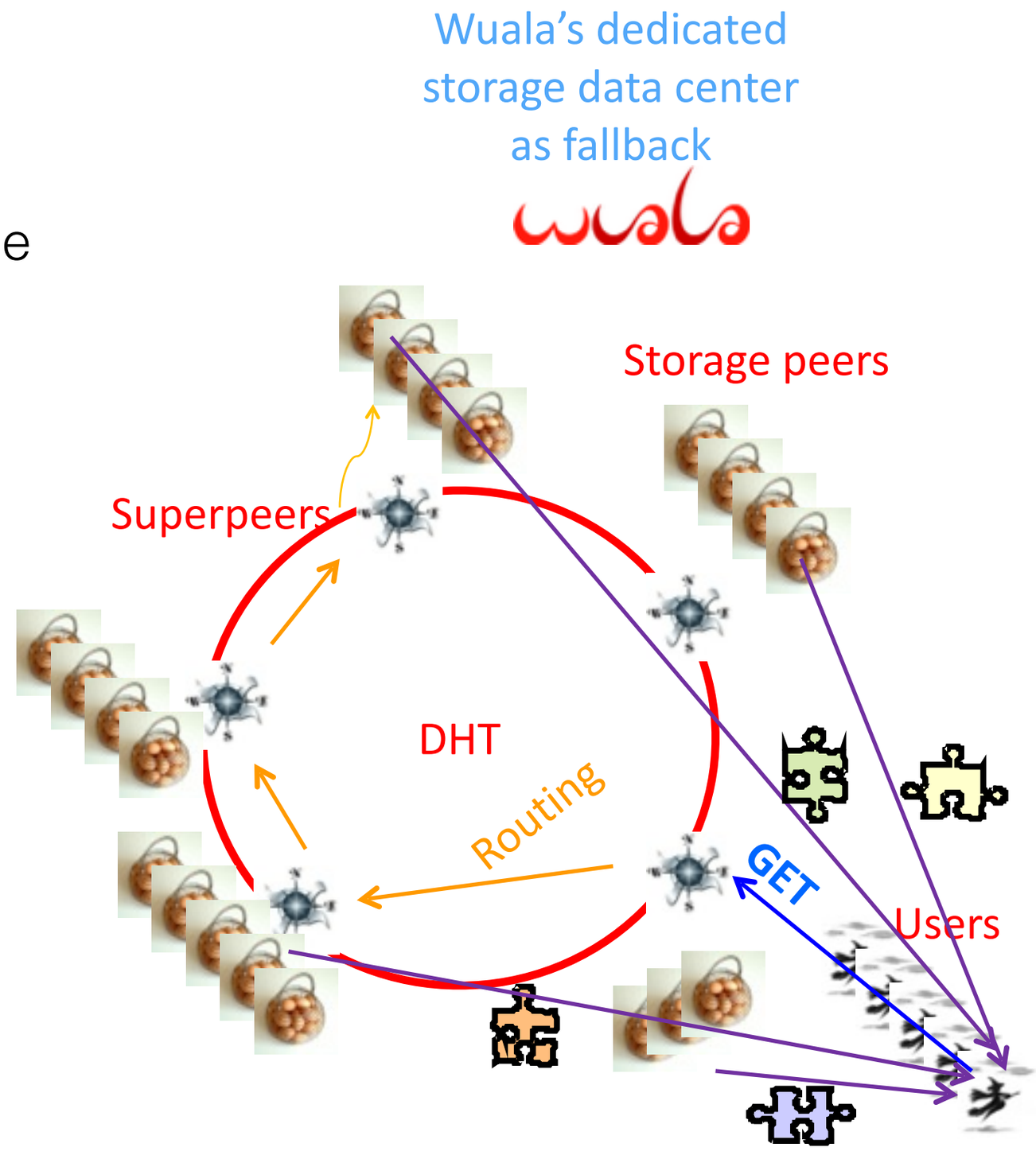
- Hybrid architecture (used previously in Wuala)



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

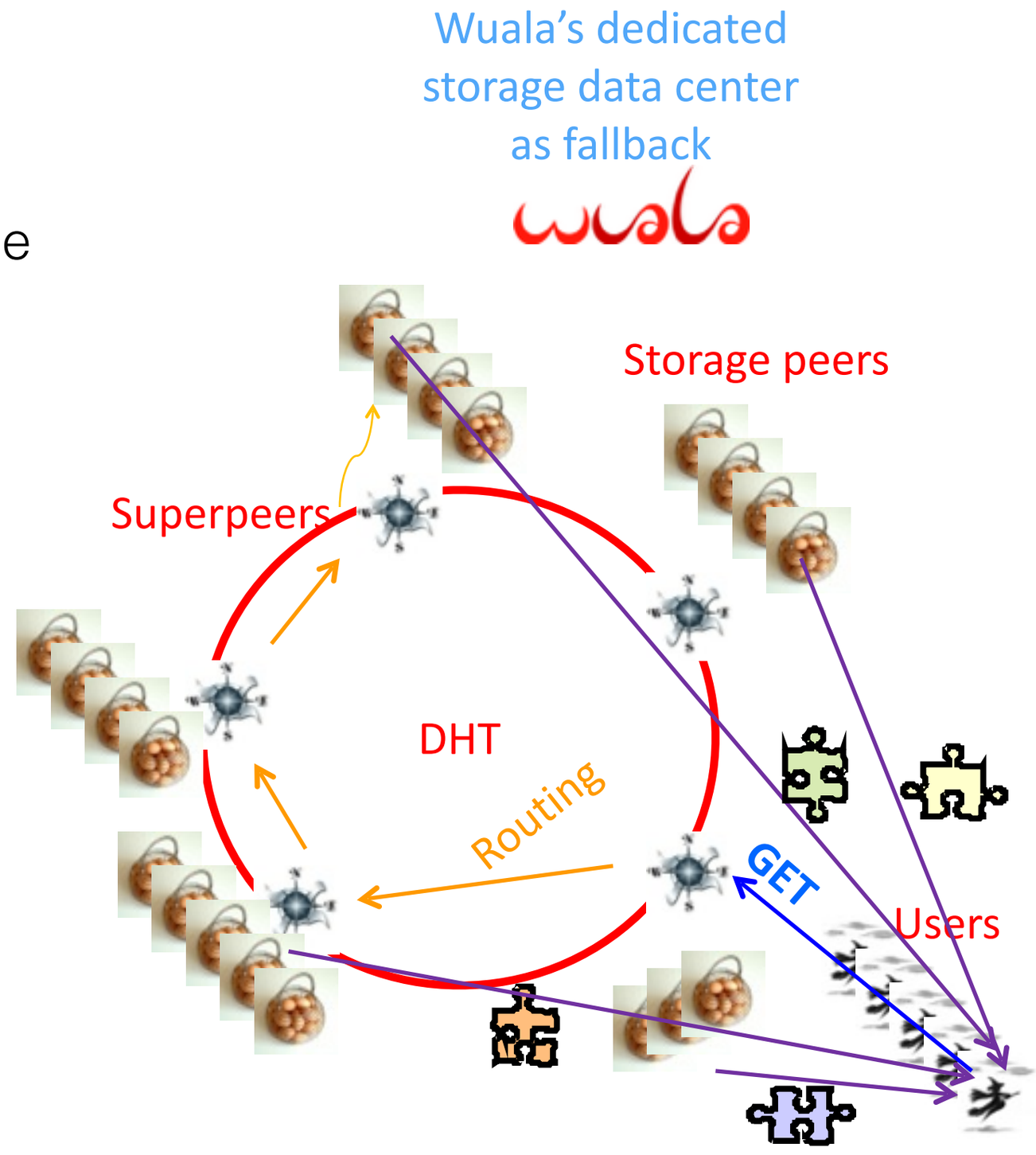
- Hybrid architecture (used previously in Wuala)
  - Index independent of storage



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

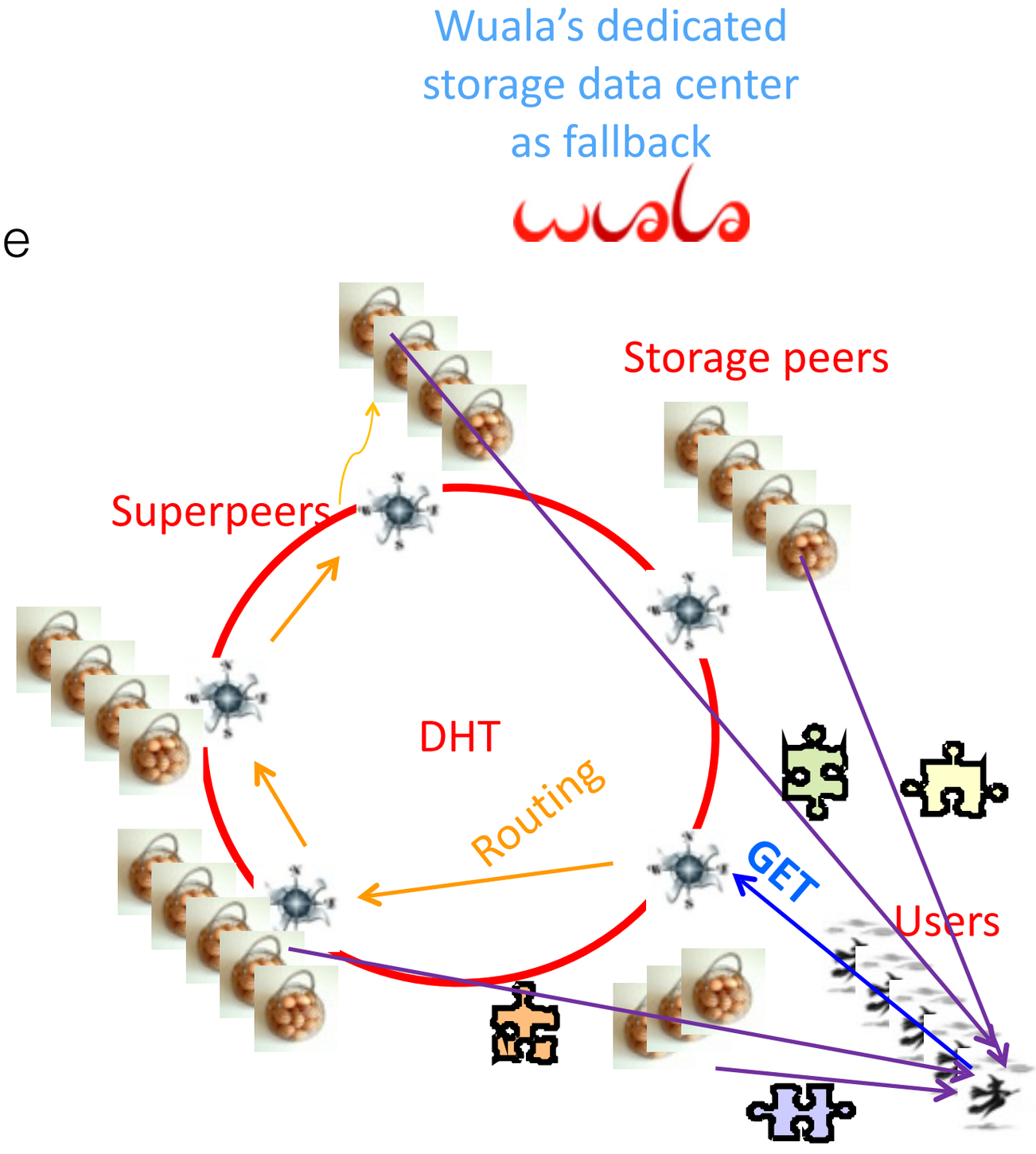
- Hybrid architecture (used previously in Wuala)
  - Index independent of storage
  - Many fragments per object



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

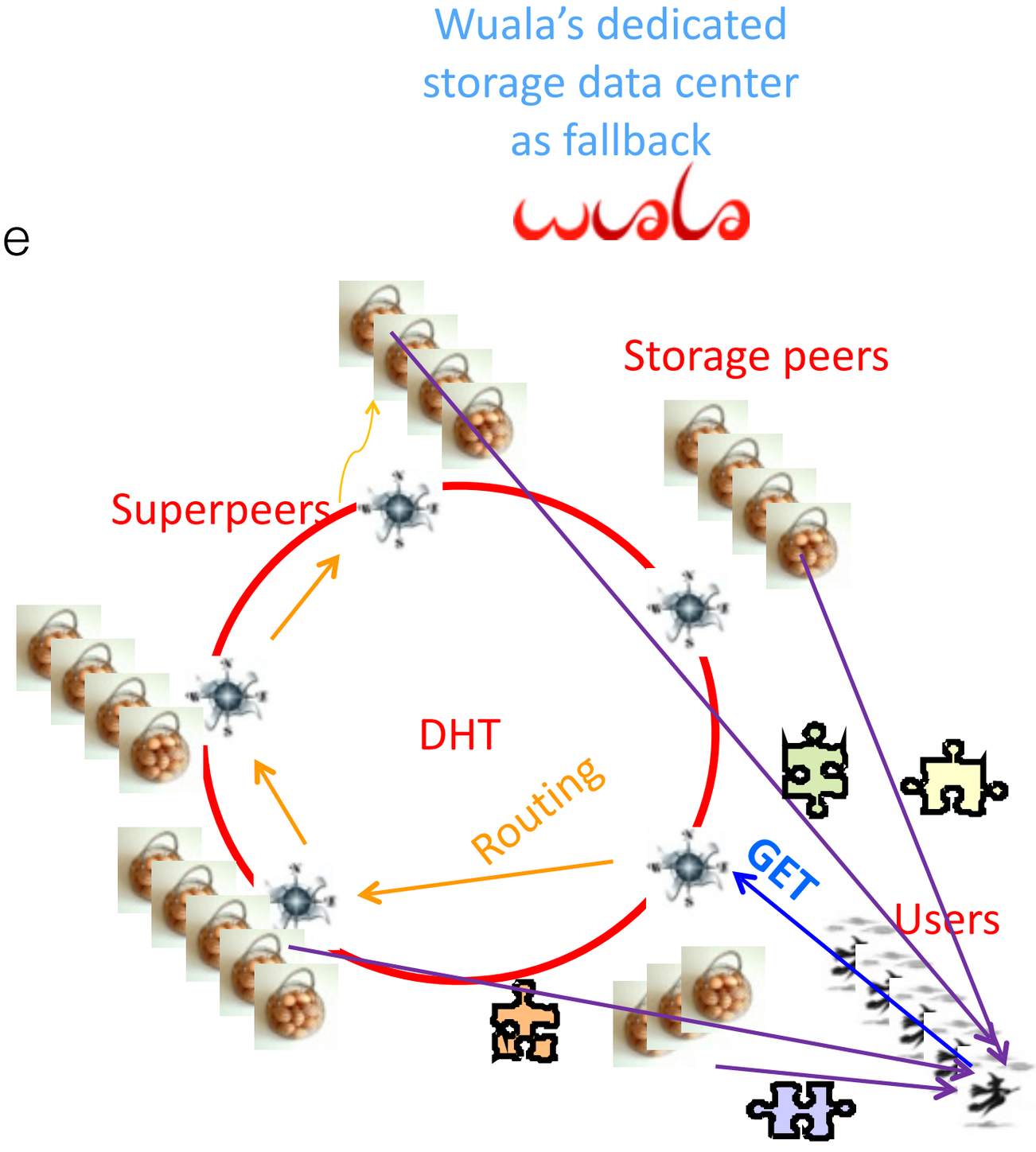
- Hybrid architecture (used previously in Wuala)
  - Index independent of storage
  - Many fragments per object
    - Suitable for sharing very large but static files



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

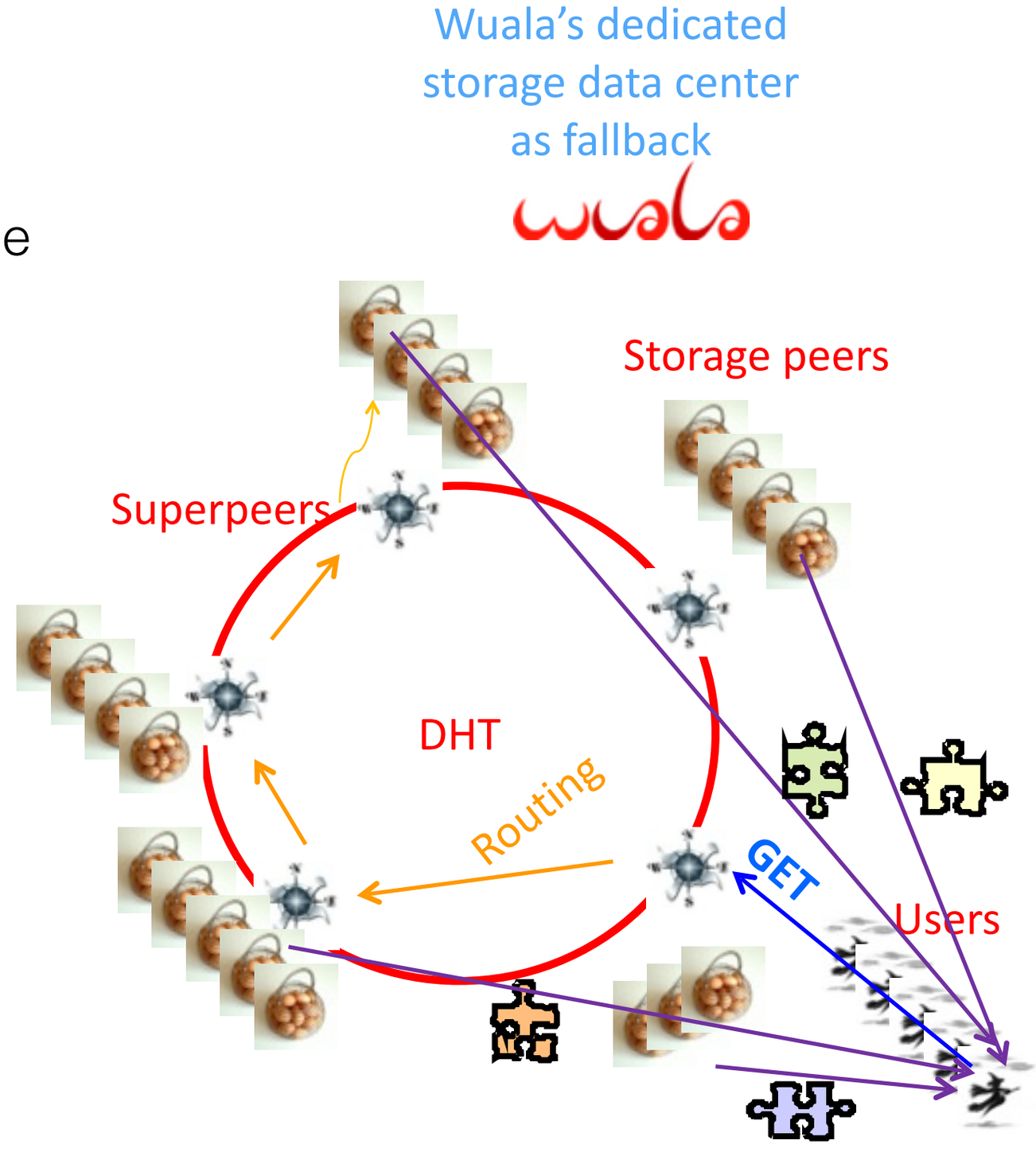
- Hybrid architecture (used previously in Wuala)
  - Index independent of storage
  - Many fragments per object
    - Suitable for sharing very large but static files
    - Parallel download



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

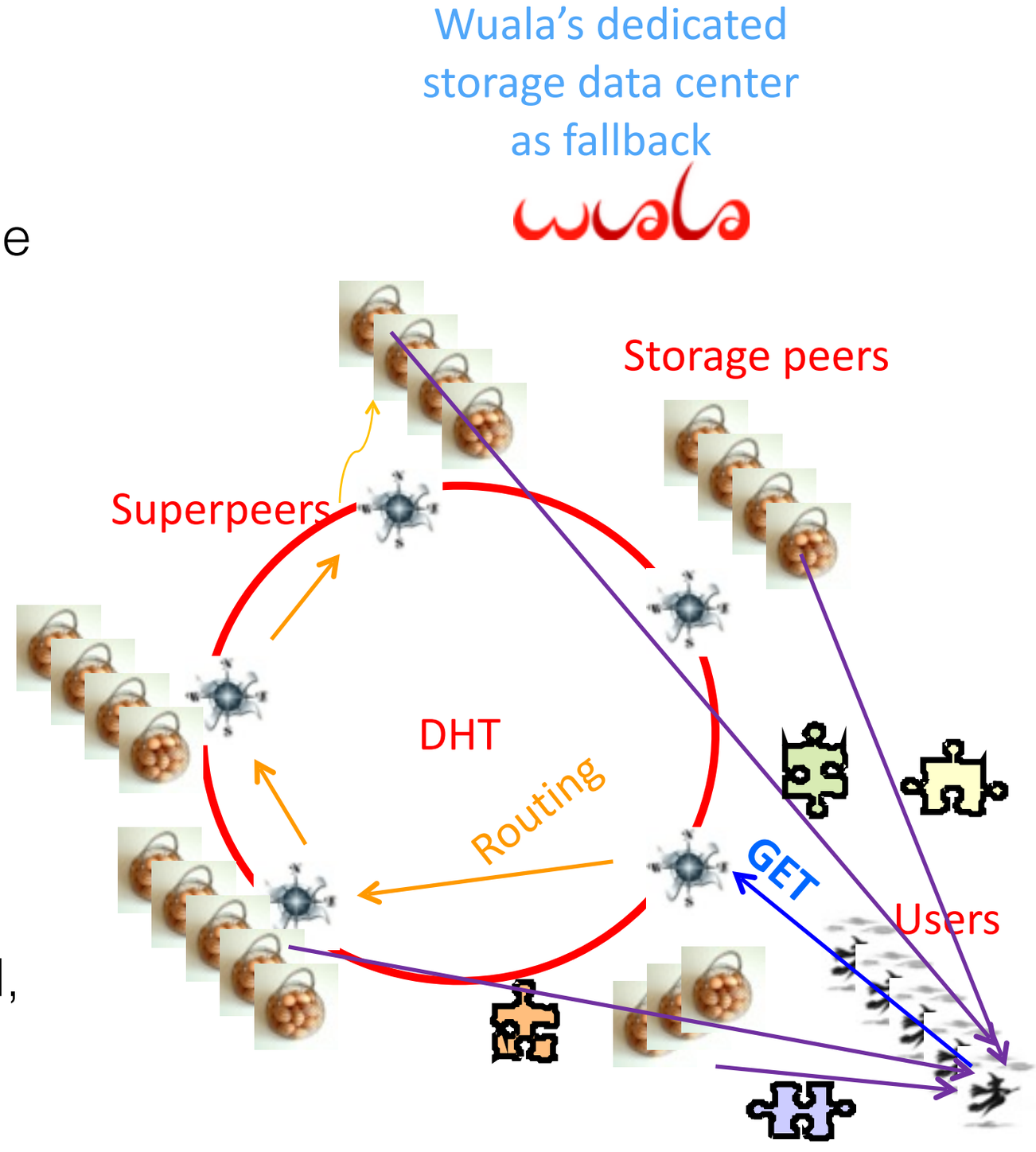
- Hybrid architecture (used previously in Wuala)
  - Index independent of storage
  - Many fragments per object
    - Suitable for sharing very large but static files
    - Parallel download
  - Piggy-backed, large DHT routing states



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

Cloud assisted storage system

- Hybrid architecture (used previously in Wuala)
  - Index independent of storage
  - Many fragments per object
    - Suitable for sharing very large but static files
    - Parallel download
  - Piggy-backed, large DHT routing states
    - So very few hops needed, gives high through-put



Source: Google tech talk on Wuala: http://www.youtube.com/watch?v=3xKZ4KGkQY8

More sophisticated heuristics

More sophisticated heuristics

- Incentives
  - reciprocity, trust/reputation, …

More sophisticated heuristics

- Incentives
  - reciprocity, trust/reputation, …
- QoS: 24/7 coverage, locality, …
  - online/offline behavior (history/prediction), …

More sophisticated heuristics

- Incentives
  - reciprocity, trust/reputation, …
- QoS: 24/7 coverage, locality, …
  - online/offline behavior (history/prediction), …
- Control
  - De/centralized, local/global knowledge

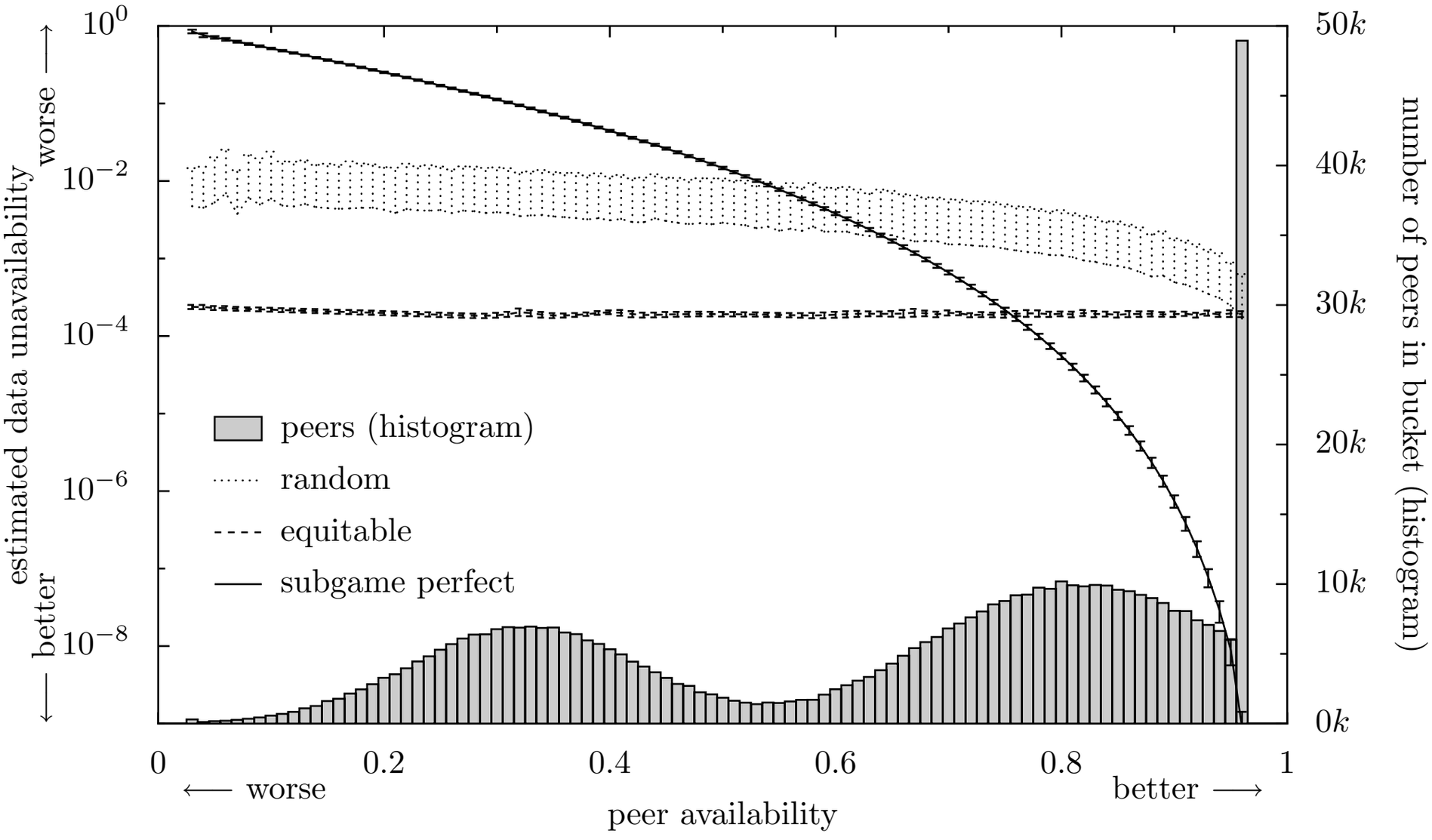
Replica Placement in P2P Storage: Complexity and Game Theoretic Analyses

Rzadca et al, ICDCS 2010

- Replication model: A clique of replicas storing each other's data (reciprocity)
  - Explores both centralized and decentralized settings for clique formation
  - Challenge
    - Centralized matching - right set of peers to optimize storage capacity utilization (proven NP-hard)
    - Decentralized matching - uses an underlying gossip algorithm (T-man) to explore partners

Representative result

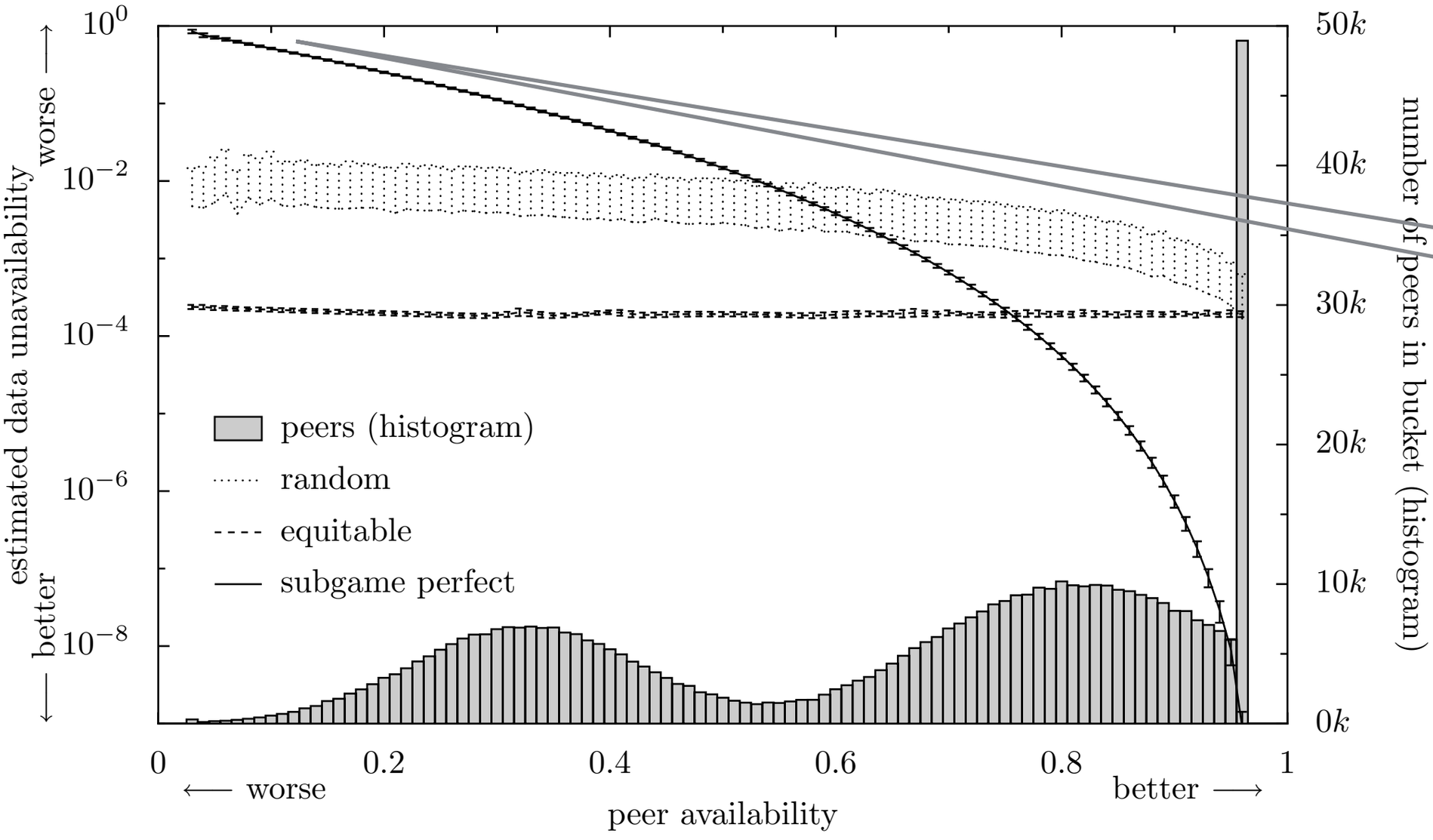
(simulations with artificial data)



Peers' expected data unavailability as a function of their availability in random, equitable and subgame perfect assignment. Histogram shows the number of peers in each availability bucket.

Representative result

(simulations with artificial data)



Good or bad?

Peers' expected data unavailability as a function of their availability in random, equitable and subgame perfect assignment. Histogram shows the number of peers in each availability bucket.

How about F2F storage?

How about F2F storage?

- Friend-to-Friend instead of Peer-to-Peer

How about F2F storage?

- Friend-to-Friend instead of Peer-to-Peer
  - Translating “real life” trust into something useful for reliable “system” design

How about F2F storage?

- Friend-to-Friend instead of Peer-to-Peer
  - Translating “real life” trust into something useful for reliable “system” design

  - Maps naturally to the overlying social application

How about F2F storage?

- Friend-to-Friend instead of Peer-to-Peer
  - Translating “real life” trust into something useful for reliable “system” design

  - Maps naturally to the overlying social application
    - Anecdotal note: SafeBook used Friend-of-Friends for access control also

Place data at friends: That's it?

- Store at all friends (naïve/baseline)
  - Best one can do in terms of achieving highest possible availability
  - Very high overheads!
    - Storage
    - Maintenance

Place data at friends: That's it?

- Store at all friends (naïve/baseline)
  - Best one can do in terms of achieving highest possible availability
- Very high overheads!
  - Storage
  - Maintenance

Find instead a "reasonable" subset of friends to store at!

An empirical study of availability in friend-to-friend storage systems

Sharma et al, P2P 2011

An empirical study of availability in friend-to-friend storage systems

Sharma et al, P2P 2011

- Look at the temporal online/offline behavior of friends

An empirical study of availability in friend-to-friend storage systems

Sharma et al, P2P 2011

- Look at the temporal online/offline behavior of friends
  - Achievable coverage
    - What best availability can be achieved?

An empirical study of availability in friend-to-friend storage systems

Sharma et al, P2P 2011

- Look at the temporal online/offline behavior of friends
  - Achievable coverage
    - What best availability can be achieved?
  - Criticality of friends
    - Which friends are indispensable?

Evaluation

- Data set
  - Italian instant messenger service
    - ✦ Pros
      - Social+Temporal characterisitcs
      - “May” reasonably reflect the online/offline behavior
    - ✦ Cons:
      - Not a p2p storage system trace
      - “small”, “incomplete” and “geographically localized”

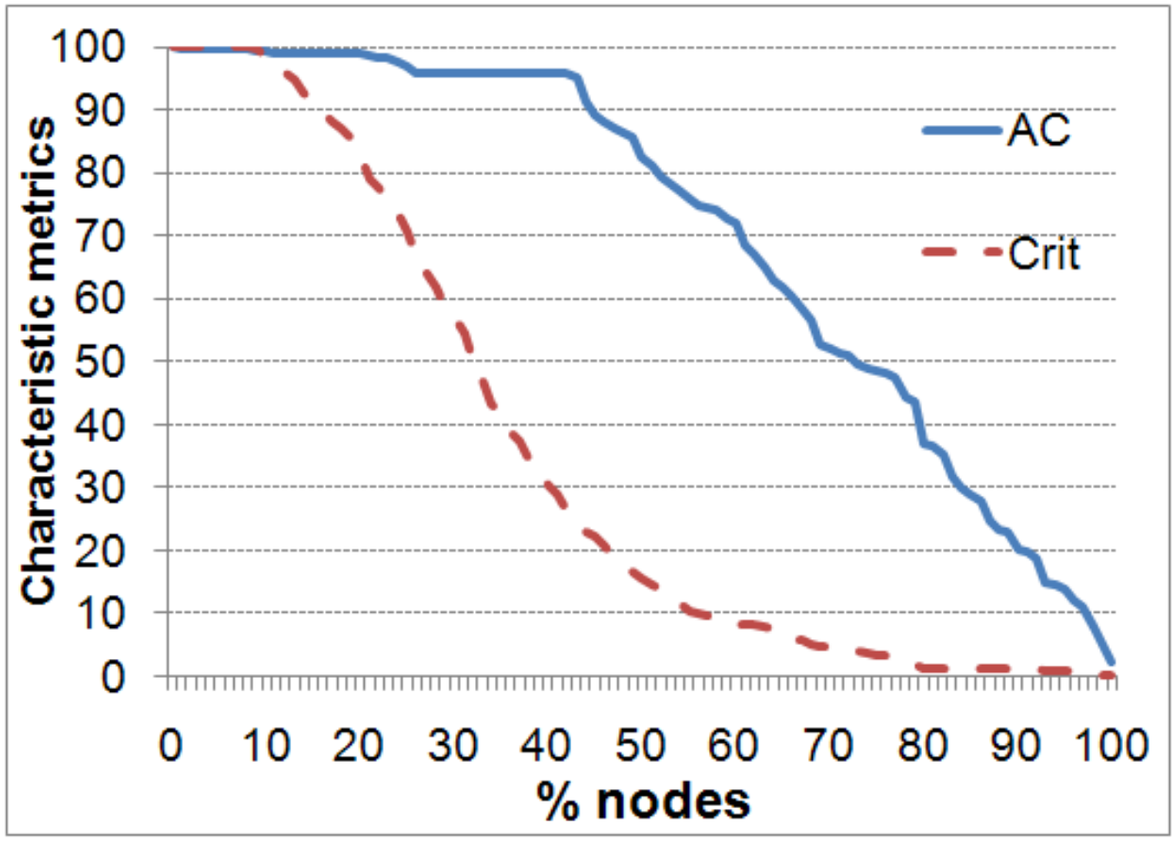
Evaluation

- Data set
  - Italian instant messenger service
    - Pros
      - Social+Temporal characterisitcs
      - “May” reasonably reflect the online/offline behavior
    - Cons:
      - Not a p2p storage system trace
      - “small”, “incomplete” and “geographically localized”

- 3436 nodes
  - 848 nodes in the largest component
    - Note that many nodes had “neighbors” in other servers, for whom we did not have info.
    - Between 1-18 neighbors
- Use two weeks of data
  - One for “learning”, one for evaluation
    - Time of day, day of week effects

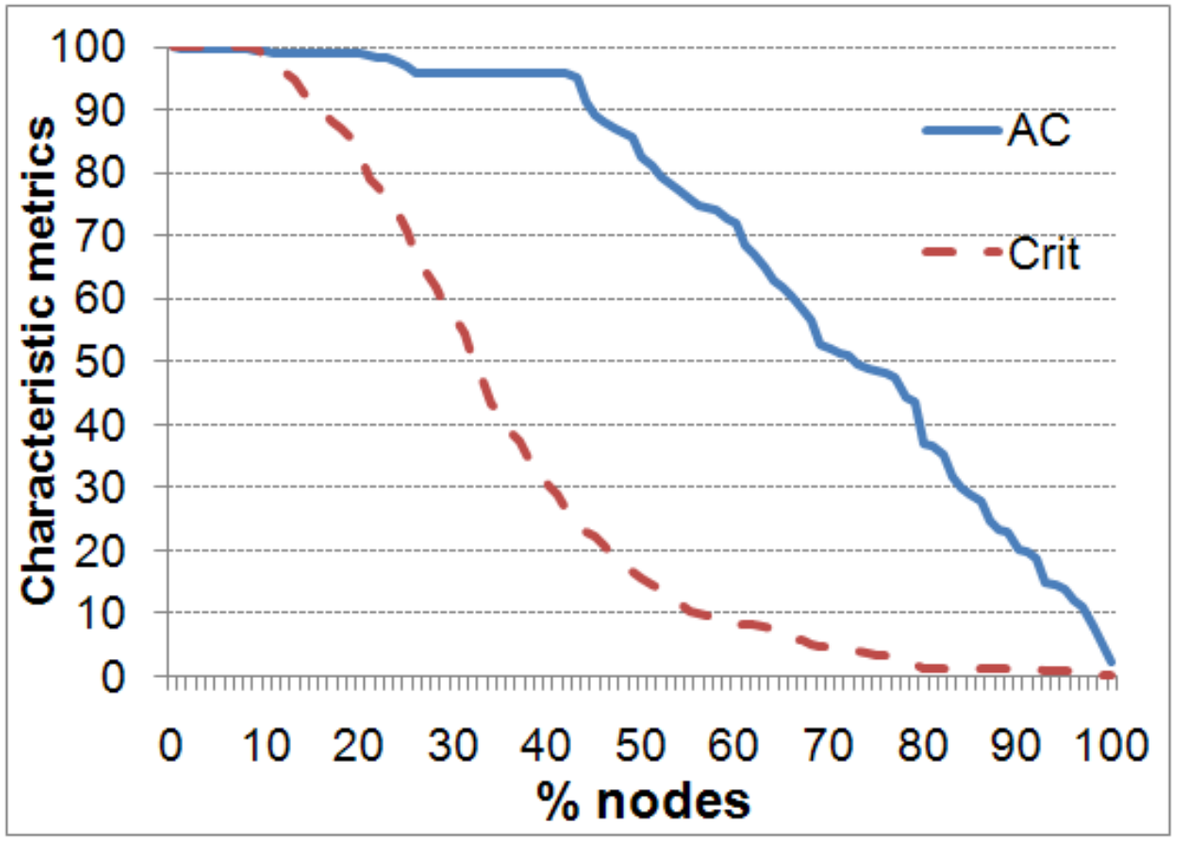
Representative results

Representative results

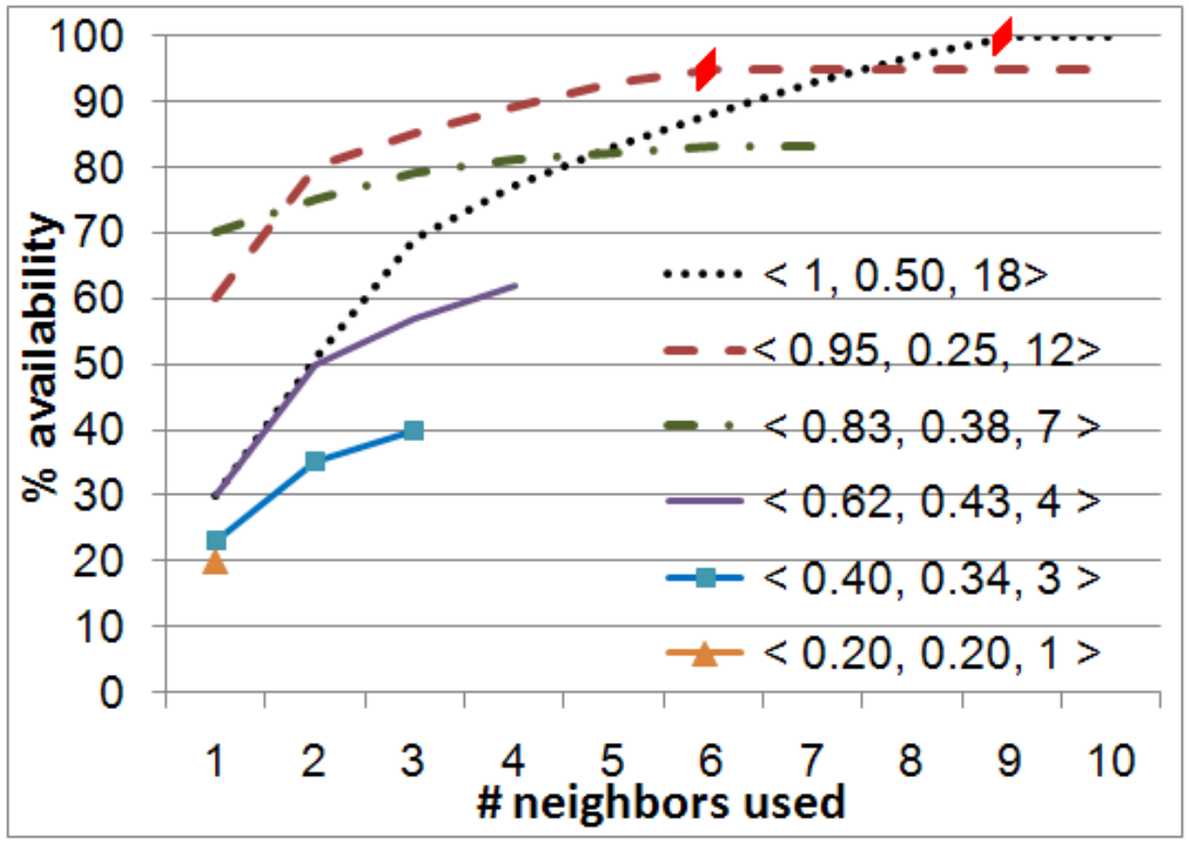


- AC: achievable coverage
  - 50% nodes can get more than 90% availability
- Crit: Time covered using critical nodes
  - Too much dependence on critical nodes

Representative results

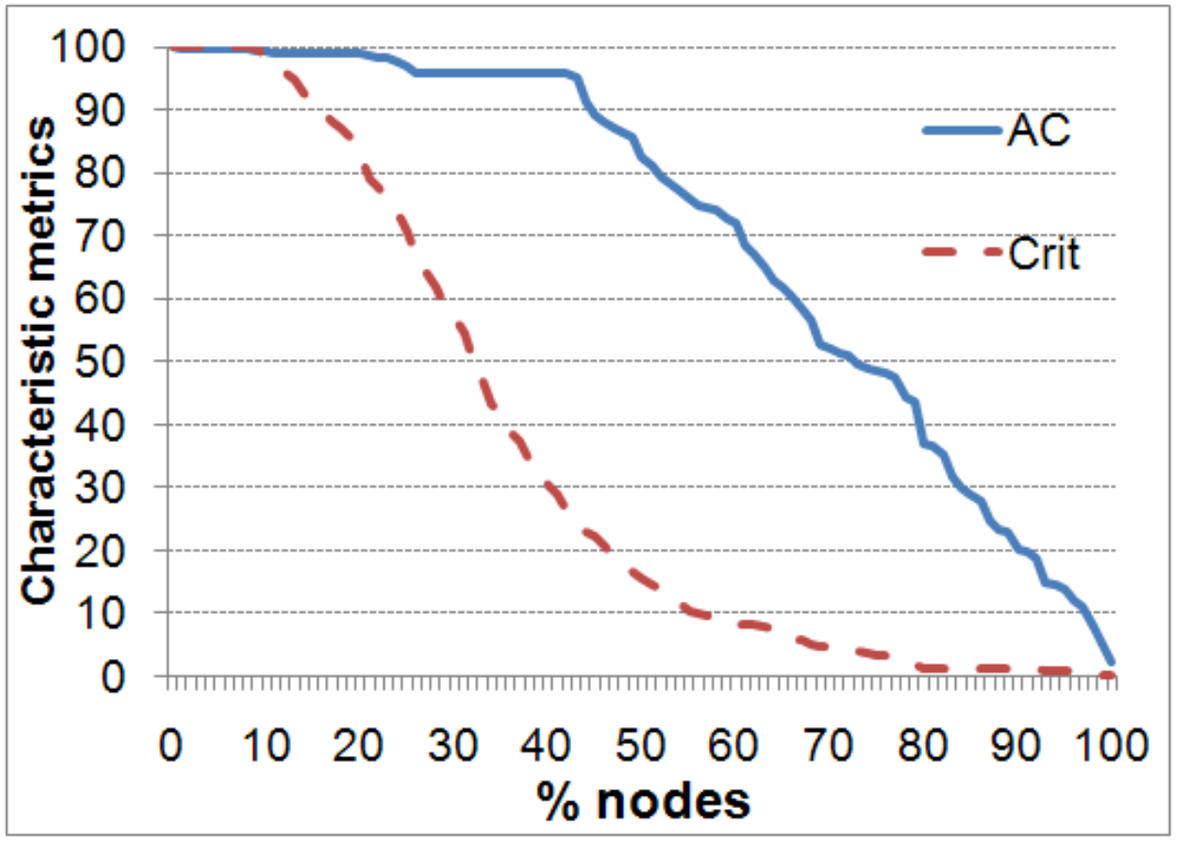


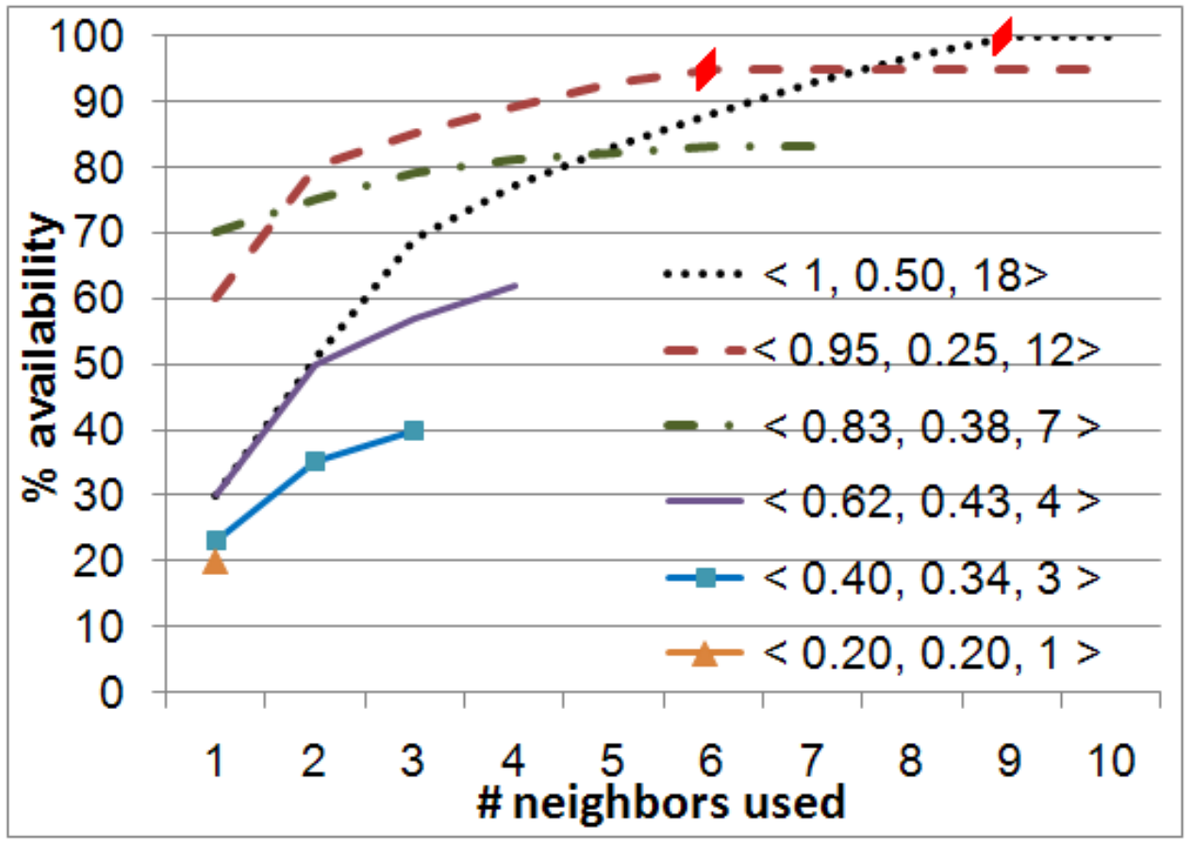
- AC: achievable coverage
  - 50% nodes can get more than 90% availability
- Crit: Time covered using critical nodes
  - Too much dependence on critical nodes



- <Achievable coverage, Degree of Criticality, # of Friends>

Representative results





- AC: achievable coverage
  - 50% nodes can get more than 90% availability
- Crit: Time covered using critical nodes
  - Too much dependence on critical nodes

- <Achievable coverage, Degree of Criticality, # of Friends>

If there are “enough” friends, (>10), ought to be okay! (assuming storage capacity is not an issue)

Bootstrapping pangs!

- New peers with few friends in the system, or no reputation of being highly available, will find it difficult to get started!
  - Game-theoretic study on reciprocity based P2P cliques
  - Analysis of ego-centric networks for F2F storage

SuperNova: Super-peers Based Architecture for Decentralized Online Social Networks

Sharma et al, Comsnets 2012

SuperNova: Super-peers Based Architecture for Decentralized Online Social Networks

Sharma et al, Comsnets 2012

- The big picture/premise

SuperNova: Super-peers Based Architecture for Decentralized Online Social Networks

Sharma et al, Comsnets 2012

- The big picture/premise
  - Well resourced nodes act as super-peers
    - incentives (could be): reputation within an interest community, ability to monetize (e.g., using ads), …

SuperNova: Super-peers Based Architecture for Decentralized Online Social Networks

Sharma et al, Comsnets 2012

- The big picture/premise
  - Well resourced nodes act as super-peers
    - incentives (could be): reputation within an interest community, ability to monetize (e.g., using ads), …
  - New nodes use superpeers for storage, until they get established in the system
    - so that the super-peers are not over-burdened, or become a bottleneck for established peers, …

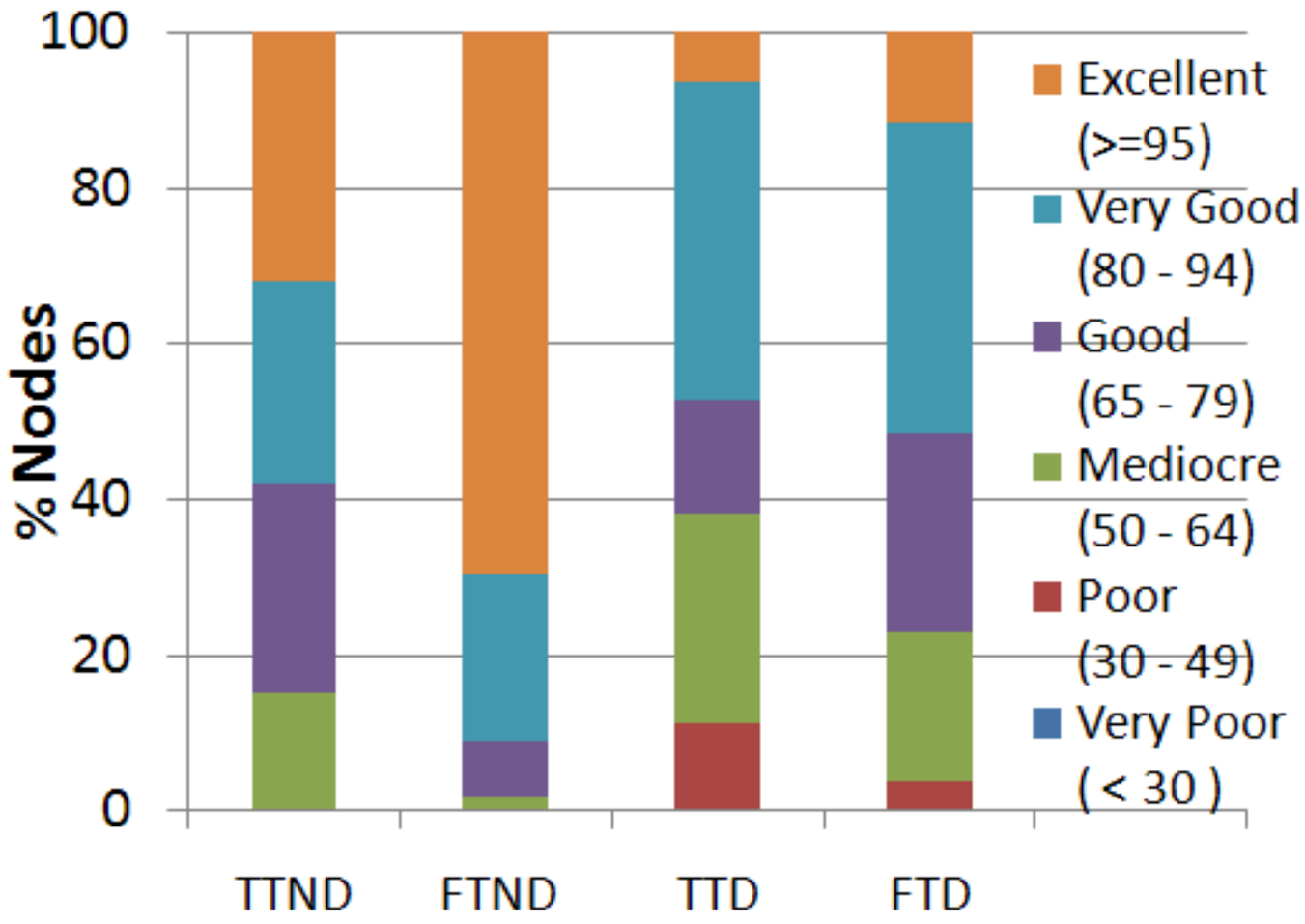
SuperNova: Super-peers Based Architecture for Decentralized Online Social Networks

Sharma et al, Comsnets 2012

- The big picture/premise
  - Well resourced nodes act as super-peers
    - incentives (could be): reputation within an interest community, ability to monetize (e.g., using ads), …
  - New nodes use superpeers for storage, until they get established in the system
    - so that the super-peers are not over-burdened, or become a bottleneck for established peers, …
  - Superpeers help coordinating, finding storage partners, etc.

Representative result

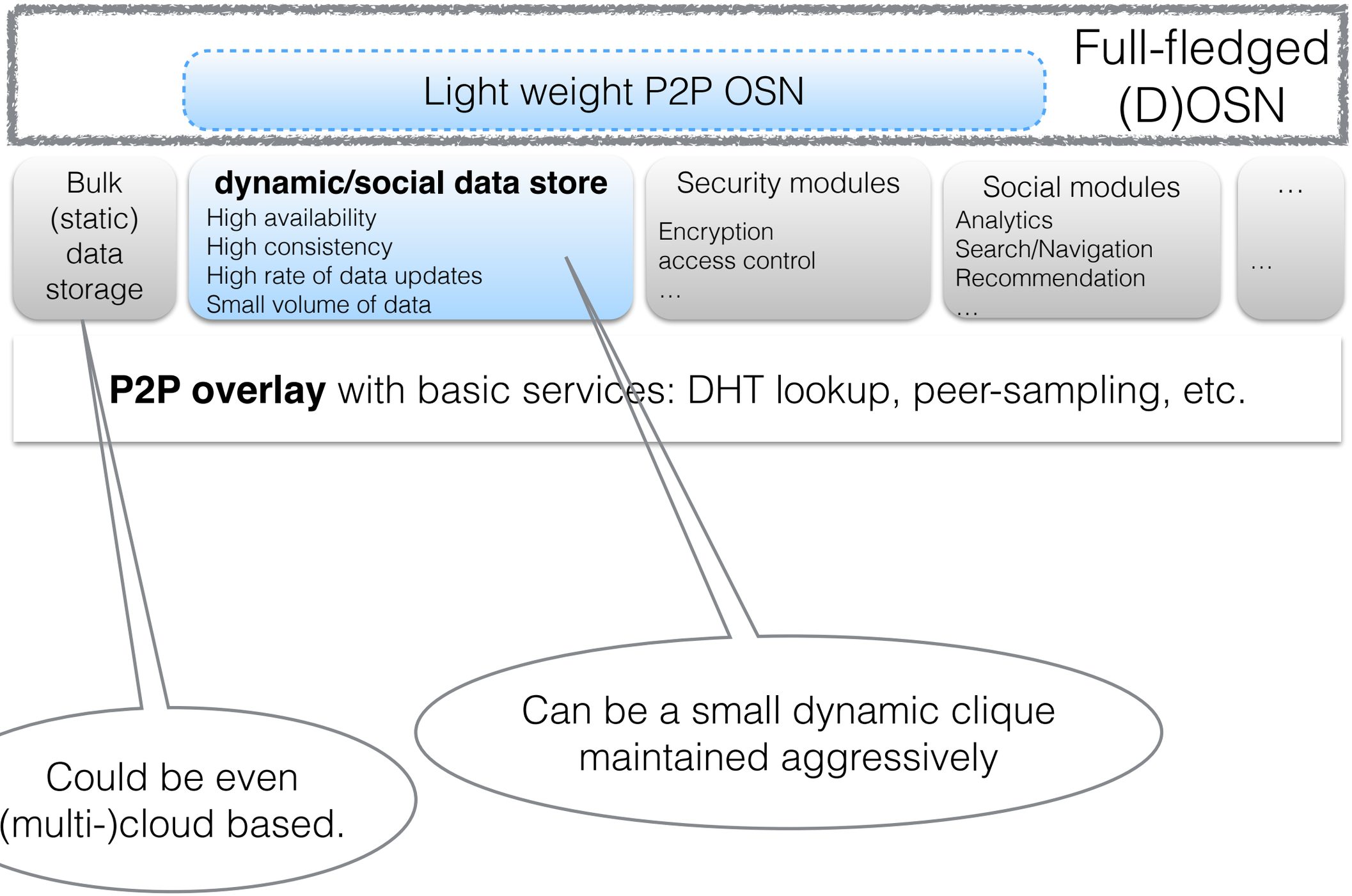
Take with a huge pinch of salt: artificial data to drive simulations, with too many parameters …



(c) System Performance

Comparison for Friend's Time (FT) and Total Time (TT) for Deviation (D) and NonDeviation (ND)

Moving forward




Tack så mycket!