



# Using the Small-World Model to Improve Freenet Performance

---

Hui Zhang   Ashish Goel  
USC

Ramesh Govindan  
ICSI



# Peer-to-peer systems

---

- IP overlay networks
  - without any central control or hierarchical organization
  - Each node in the network is equivalent in functionality
- A special class: content-addressed networks
  - a document is accessed using a descriptive title of the content rather than the location where the document is stored.
    - A data object is represented by a point in a key space.
  - At the core lies the distributed algorithm for content lookup and dissemination



# Structured vs. Unstructured

---

## ■ Structured system

- there is global consensus on which network node a document is stored.
  - Algorithmic mapping between document key and node identifier.
- Example
  - CAN, CHORD, Tapestry, Pastry.

## ■ Unstructured system

- no such consensus exists
  - Document key and node identifier are irrelevant.
- Example
  - Freenet



## Structured vs. Unstructured (cont'd)

---

### ■ Structured system

- Advantage
  - Efficient query
  - Guarantee retrieval of existing documents
- Disadvantage
  - Hard to provide anonymity
  - Problem caused by DoS and selective attacks

### ■ Unstructured system

- Advantage
  - Simplicity in network management
  - Resistant to attacks
- Disadvantage
  - Retrieval cost not bounded
  - Search may fail even if the requested data exists



# Freenet

[Clark et al 2001]

---

- A Distributed Anonymous Information Storage and Retrieval System
- An adaptive Peer-to-Peer network application
- No broadcast search or centralized location index is employed in Freenet



## Freenet routing protocol [Clark et al 2001]

---

- Files are identified by binary file keys obtained by applying a hash function.
- Each node maintains a datastore and a routing table of <key,pointer> values
- Steepest-ascent hill-climbing search with backtracking.
- LRU (Least Recently Used) route-cache Replacement Scheme

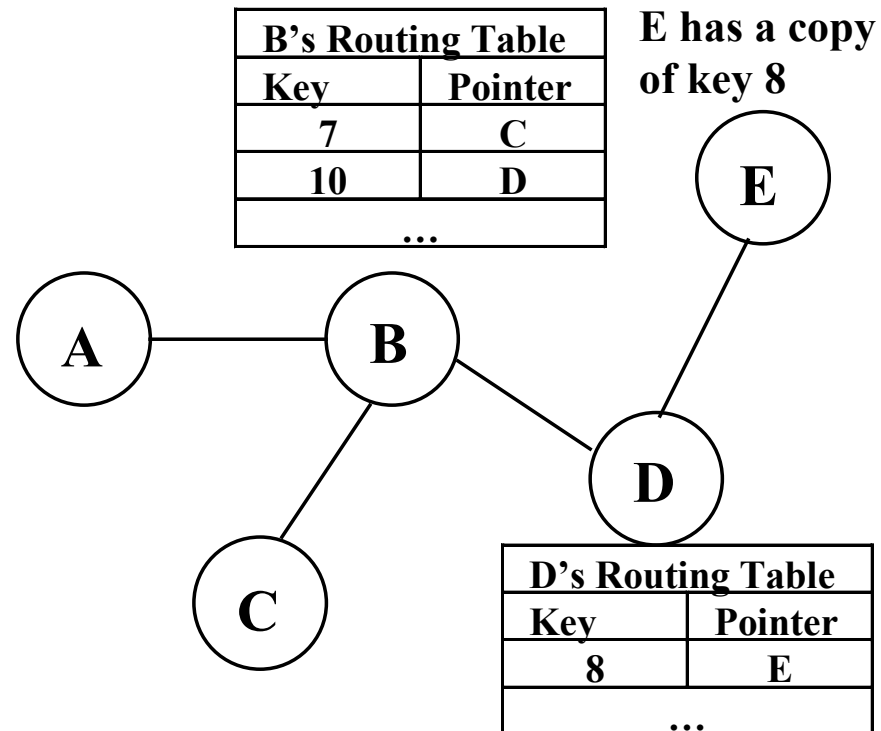


## Our contributions

---

- A new cache replacement scheme for Freenet
  - A small change to local user behavior leads to a significant global improvement on system performance.
- **Discovery:** A carefully configured unstructured system gives comparable performance as structured systems in terms of the size of the routing table and the average number of hops per request.

# An example search in Freenet network





## Main premise behind Freenet

---

- The routing in Freenet is expected to run efficiently
  - nodes should come to specialize in locating sets of similar keys
  - nodes should become similarly specialized in storing clusters of files having similar keys.

# Simulation Results (I)

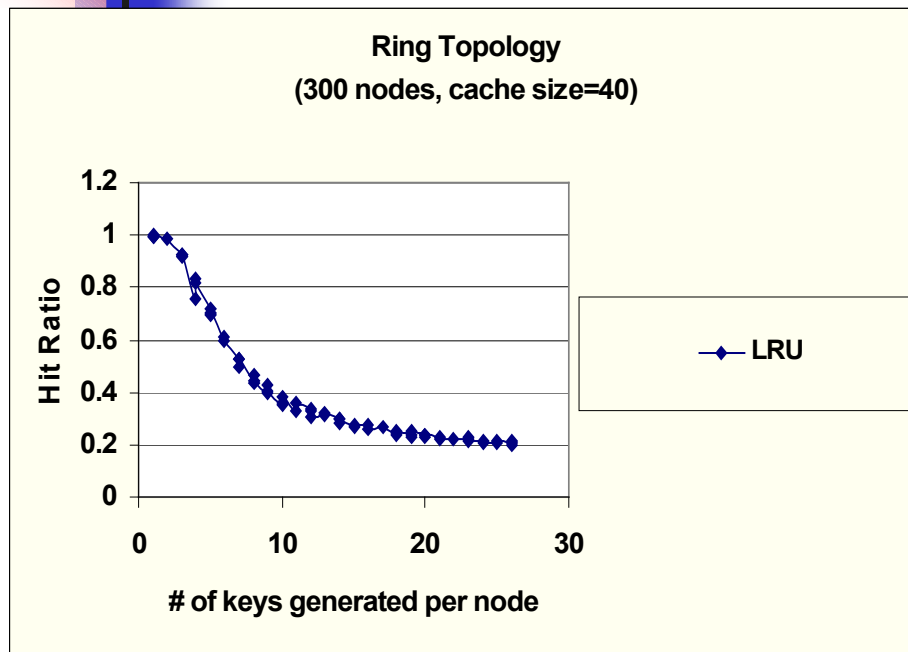


Fig. 1: Hit Ratio vs. Key Size

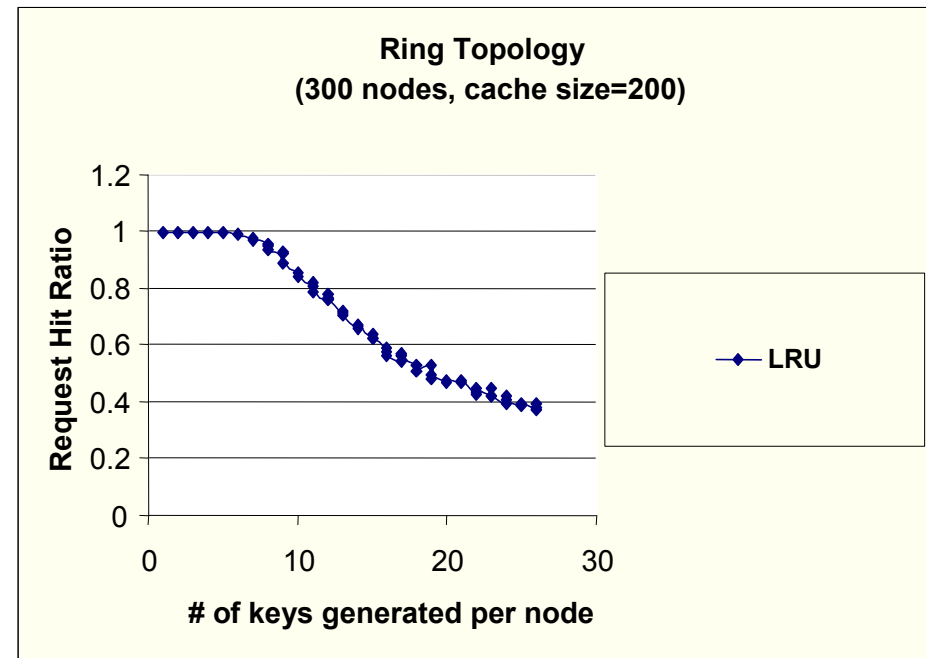


Fig.2: Simulation with a larger datastore and routing table.

For the **original** Freenet protocol, there is a **steep reduction** in the **hit ratio** with increasing load .



## Why can't Freenet search efficiently under heavy work load?

---

- Hypothesis: Frequent local caching actions could break up clusters caused by the Freenet routing mechanism
- Our simulation results showed that Freenet does not achieve clustering with light randomness in its route cache when lots of data items were inserted into the network

# Snapshots of Freenet node's datastore

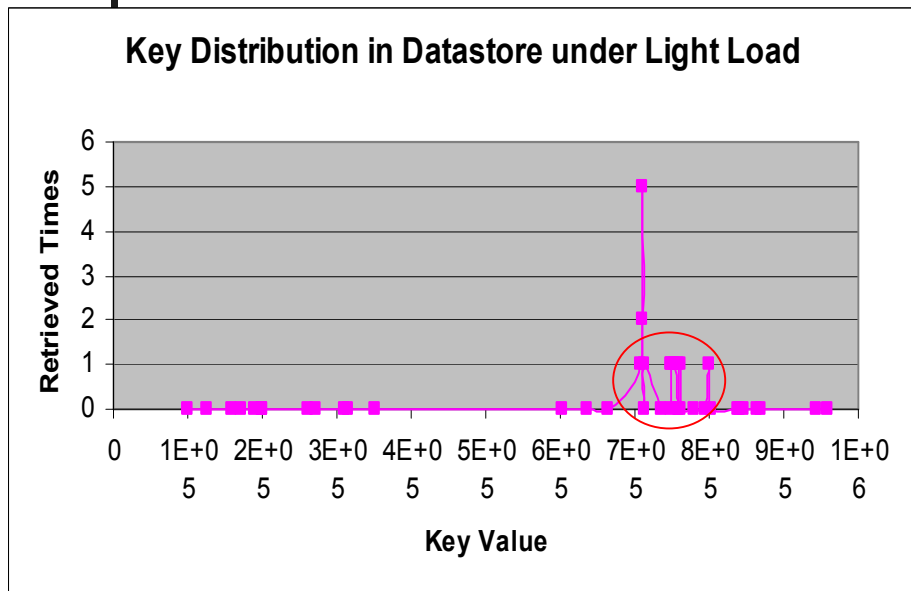


Fig.3: average number of keys generated per node =2

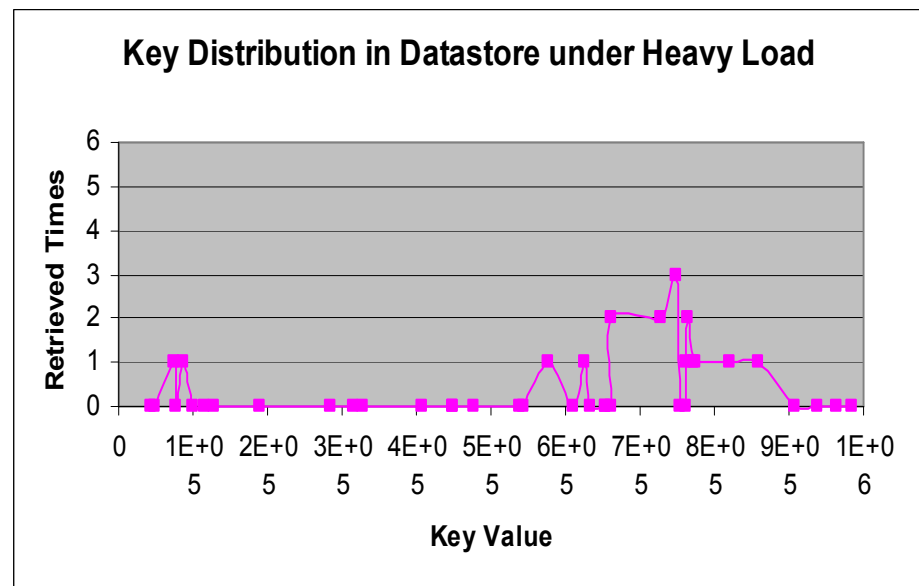


Fig.4: average number of keys generated per node =20



# Problem Definition

---

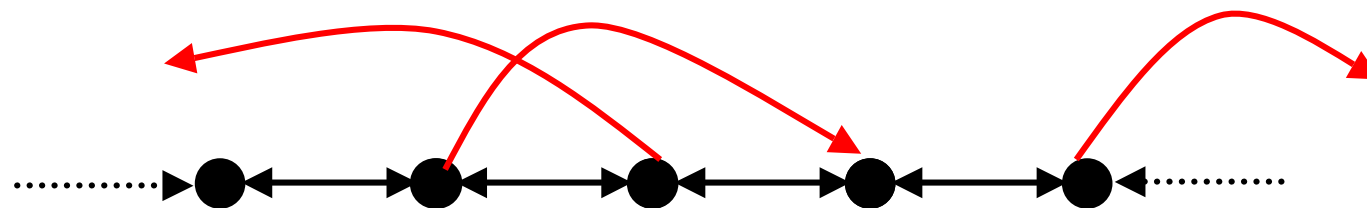
- How can we improve the routing performance of the Freenet protocol efficiently without affecting its design goals?
- Two obvious solutions which do not work
  - Increasing the cache-size
    - the cache-size is locally administered since it depends on individual system resources.
  - Increasing the HopsToLive value
    - could increase the hit ratio, at the expense of significantly increased access latency.

# Small-world model [Watts et al 1998]

- A network between order and randomness
  - short-distance clustering (like regular graph)
  - + long-distance shortcuts (result in short global path length like random graph)

→ Local contact

→ Shortcut



An one-dimensional small-world network example



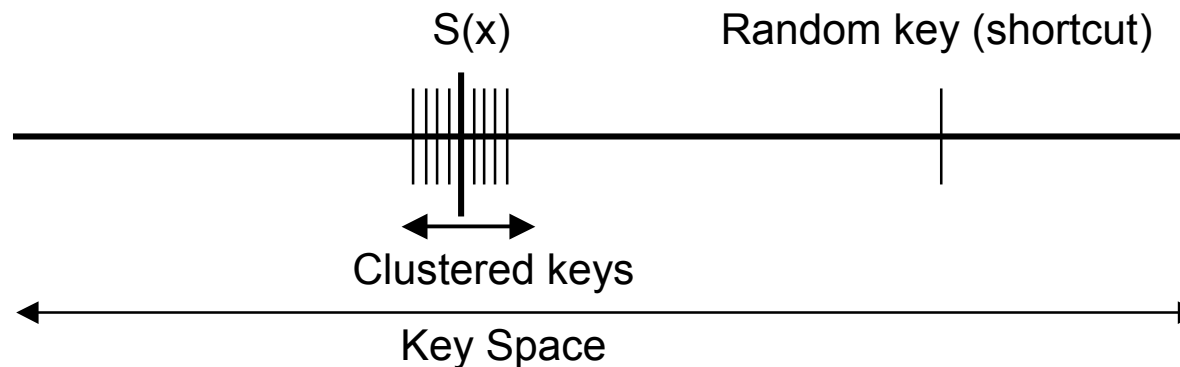
## Kleinberg's theorem [Kleinberg 1999]

---

- The network model consists of a one-dimensional linear network plus one long-distance shortcut for each node.
- The expected steps to delivery a message in the network model is  $O(\log^2 n)$  when the shortcut for each node is chosen with the probability inversely proportional to the distance.

# Clustering with randomness

- To improve routing performance, we want the routing table at node  $x$  to conform to the small-world model. It needs a set of key entries clustered around some key  $s(x)$  and one or more randomly chosen shortcut keys



- **Crucial Observation:** Such clustering can be achieved by just changing the route-cache replacement policy



## Enhanced-clustering cache replacement scheme

---

- Each node chooses a seed randomly when joining the network
- When a new key (file) **u** is to be cached, the node chooses in the current datastore the key **v** farthest from the seed

**Distance (v, seed) = Max Distance (x, seed)**

- If Distance (u, seed) < Distance (v, seed), cache **u** and evict **v** (clustering)
- If Distance (u, seed) > Distance (v, seed), cache **u** and evict **v** with probability **P** (randomness)
  - Can make **P** dependent on the two distances

# Simulation Results (II)

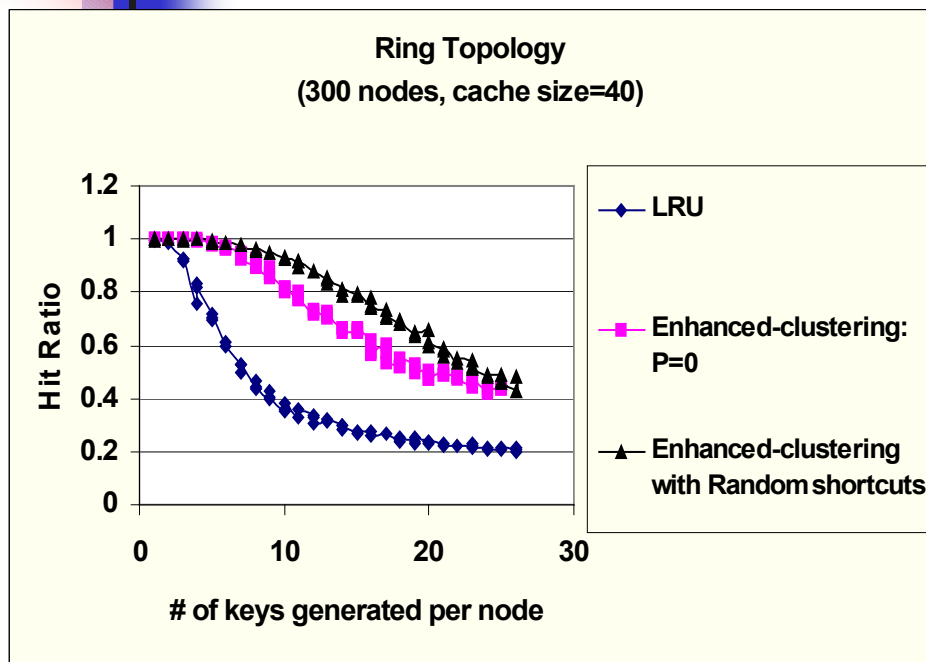


Fig. 3: Hit Ratio vs. Key Size

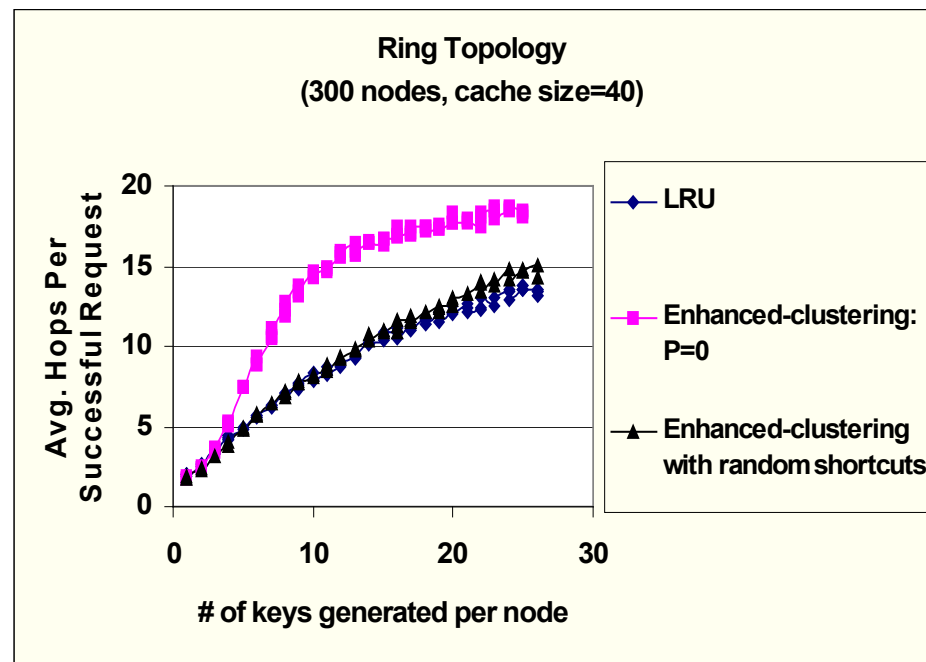
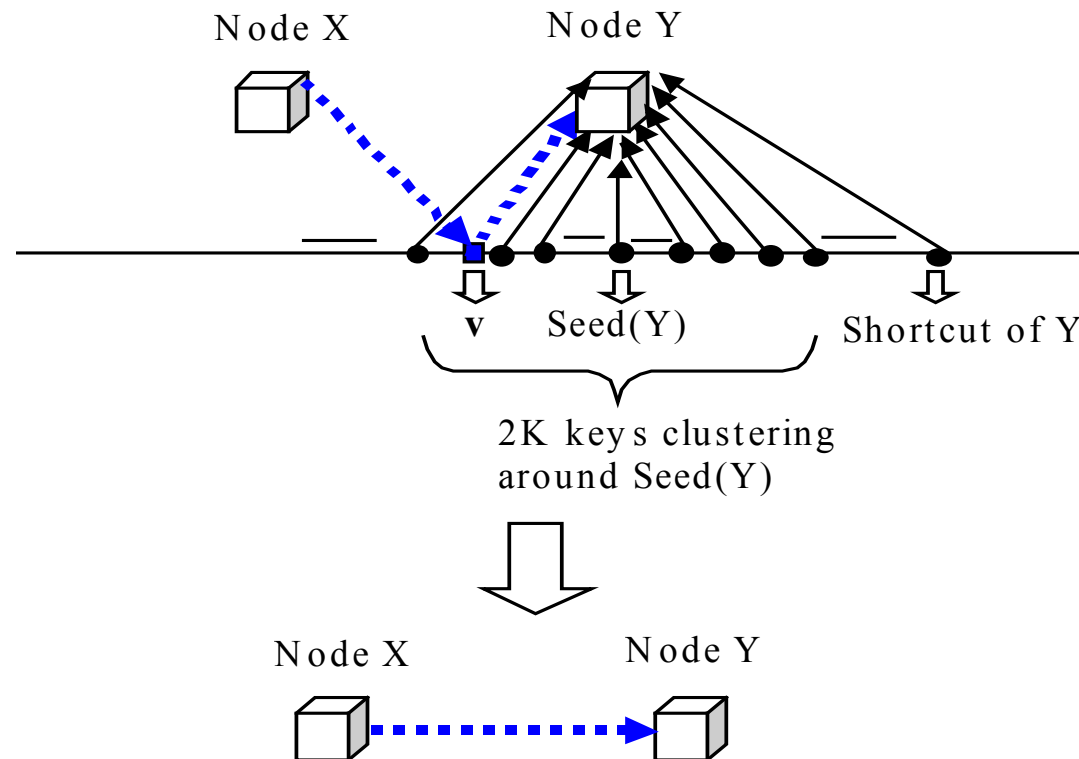


Fig. 4: Avg. Hops per Successful Request vs. Key Size

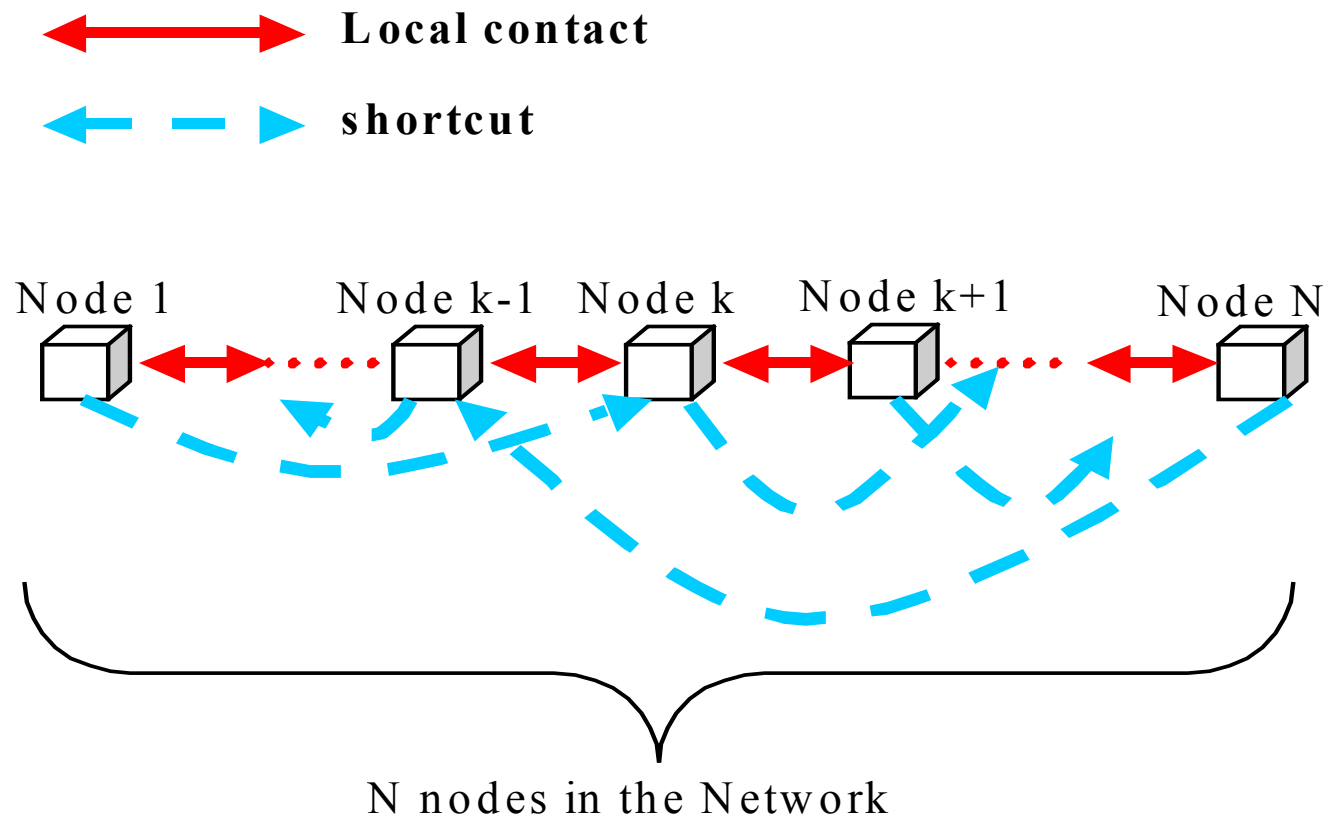
Enhanced-clustering with random shortcuts achieves both the **highest** hit ratio and the **lowest** avg. hops per successful request.

# Network model: transit keys link nodes to nodes





## An idealized model of Freenet from the node level





# Theorem 1

---

## In the idealized network model

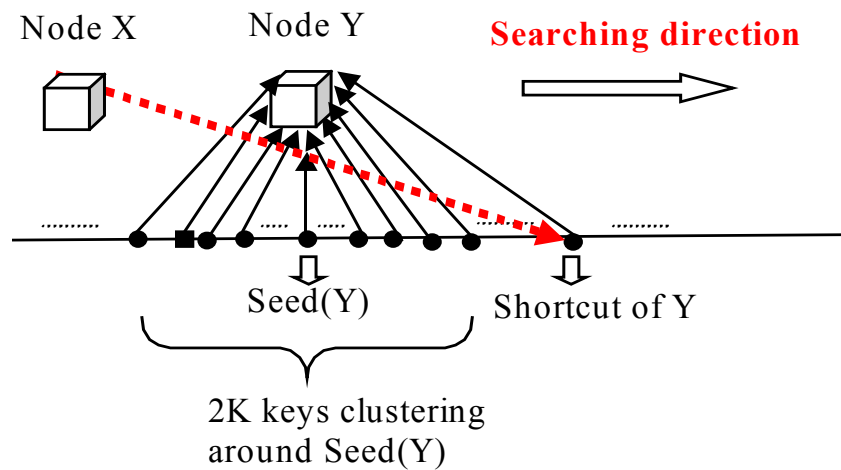
if each node  $x$  chooses its random shortcut so that the random shortcut has an endpoint  $y$  with probability proportional to  $1/d$  where  $d = |s_x - s_y|$



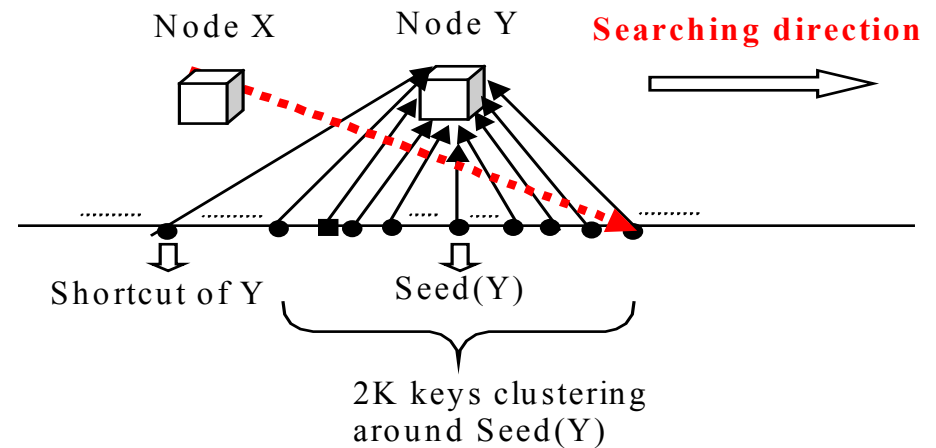
$E$  [Number of hops per request] =  $O(\log^2 n)$

using Freenet's search algorithm.

# “misleading” links: two cases



(a) A request is passed from X to Y through the shortcut of Y



(b) A request is passed from X to Y through the rightmost key of Y



## Theorem 2

---

**In the idealized network model considering the effect of “misleading” links**

if the routing table size is  $\theta(\log^2 n)$  and the shortcuts are chosen in the same way as in Theorem.1



E [Number of hops per request] =  $O(\log^2 n)$

using Freenet's search algorithm.



# Conclusion

## Performance Comparison of several P2P systems

System	Expected hops per Request	Expected Routing table size
<b>CAN</b> [Ratnaswamy et al 2001]	$O(dn^{1/d})$	$O(d)$
<b>CHORD</b> [Stoica et al 2001]	$O(\log n)$	$O(\log n)$
<b>Tapestry</b> [Zhao et al 2000]	$O(\log n)$	$O(\log n)$
<b>Kleingbor's unique Small-world model</b>	$O(\log^2 n)$	$O(1)$
<b>Freenet in the idealized model</b>	$O(\log^2 n)$	$O(\log^2 n)$