Double Rulings for Information Brokerage in Sensor Networks

Rik Sarkar Xianjin Zhu Jie Gao

STONY BROOK UNIVERSITY

MobiCom'06

Problem: Information Brokerage

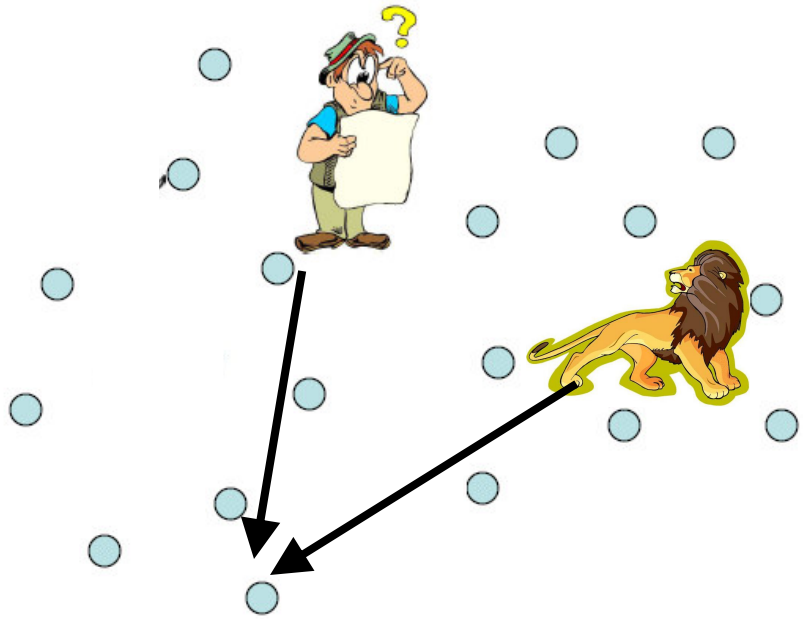
- Information producers and information consumers need to find each-other.
  - Tourists in a park looking for animals, but sensors with animals in range do not know where tourists are.

- Challenges
  - Content based search
  - Spatial/Temporal separation
  - Limited network resources

- Easy solution : Flood
  - Inefficient

Existing work: Geographic Hash Tables

- Data centric hashing.
  - Hashed node forms rendezvous
  - Enables brokerage



- Pros
  - Simple, works without flooding
- Cons
  - Nodes near hashed location become bottleneck.
  - Not distance-sensitive. Nearby producer and consumer may hash to far away nodes.

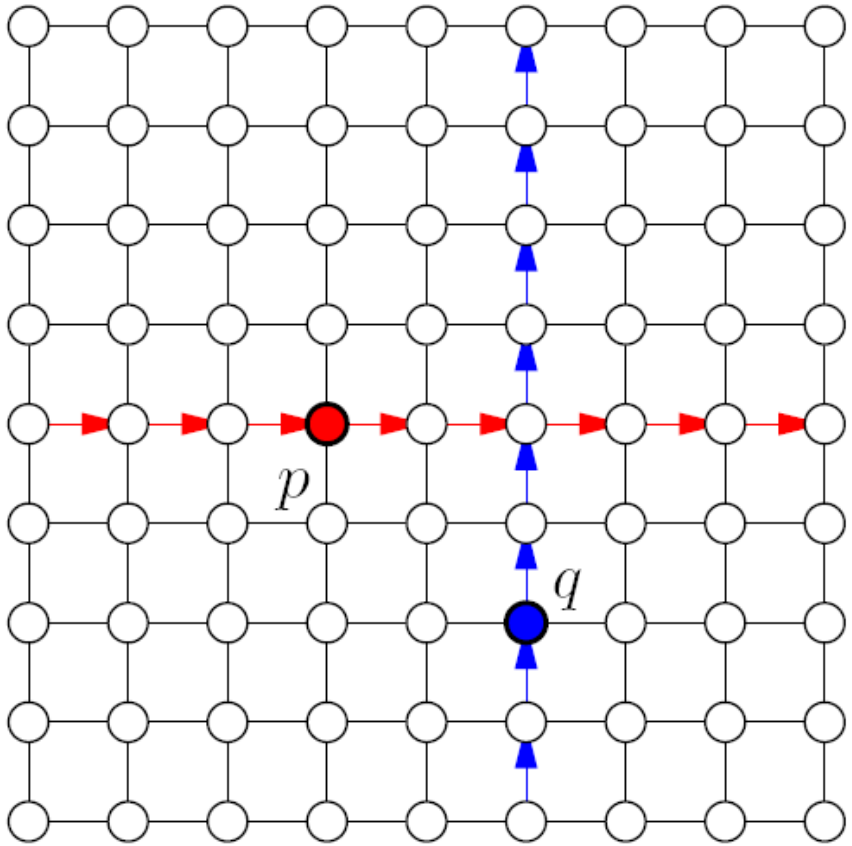
Reference: GHT [Ratnasamy, et.al. 2002].

Our Approach: Double Rulings

- Hash data to a 1-d curve, instead of a 0-d point
- Motivations for generalization
  - Data delivery uses multi-hop routing
    - Leave information along route at no extra cost
  - More flexible data retrieval
    - Easier to encounter a 1-d curve than a 0-d point

Simple Double Ruling

- Rectilinear Double Ruling
  - Producer stores data on horizontal lines
  - Consumer searches along vertical lines
  - Correctness : Every horizontal line intersects every vertical line



References: [Liu Huang Zhang 04], Rumor routing [Barginsky-Estrin 02], Quorum-based routing [Stojmenovic99].

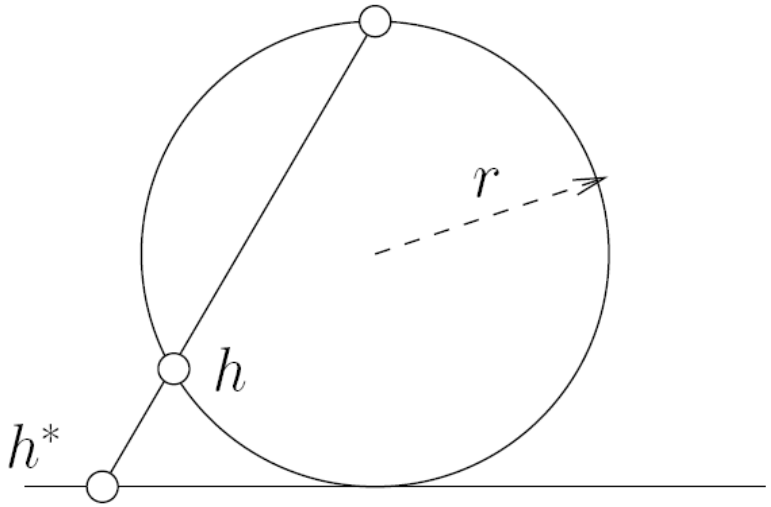
This Paper: A New Double Rulings Scheme

- Producer follows a circle to the hashed location
  - Includes GHT as a sub-case
  - Allows a large variety of retrieval mechanisms

- Improves on GHT
  - Load balancing for popular data types
  - Distance sensitivity
  - Flexible data retrieval schemes improve system robustness

Double Rulings on a Sphere

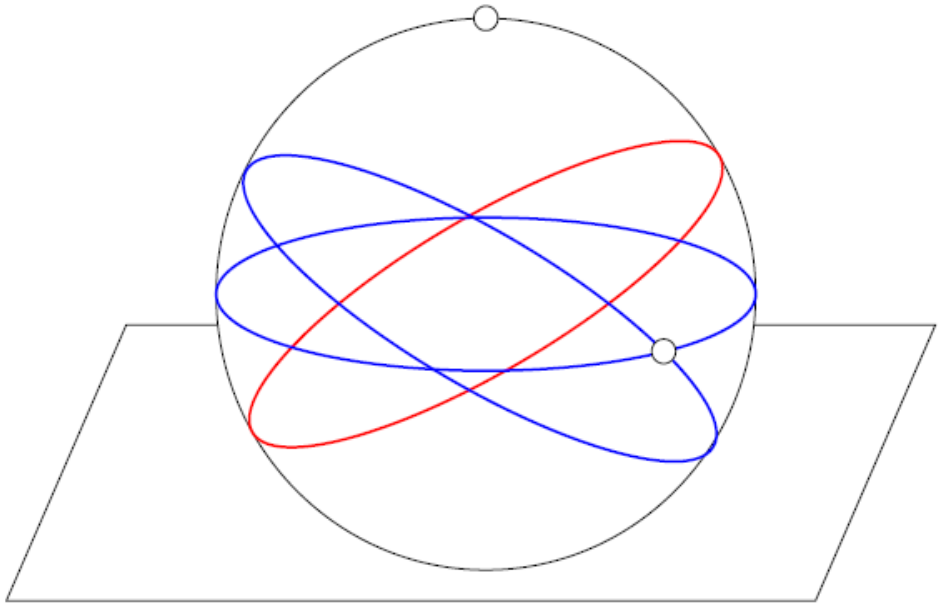
- Stereographic projection maps a projective plane to a sphere
  - Circles map to circles
  - May incur distortion

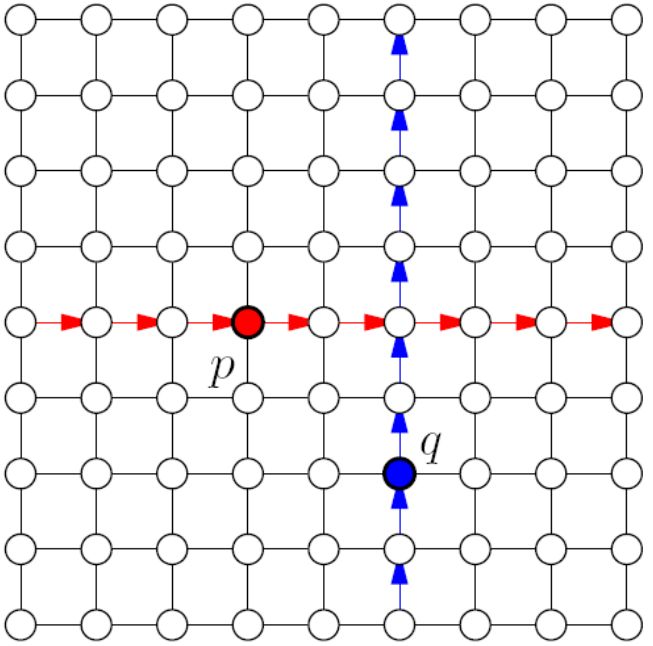


- For a finite sensor field
  - Can choose location and size of sphere such that distance distortion is bounded by $1+\varepsilon$.

Spherical Double Rulings

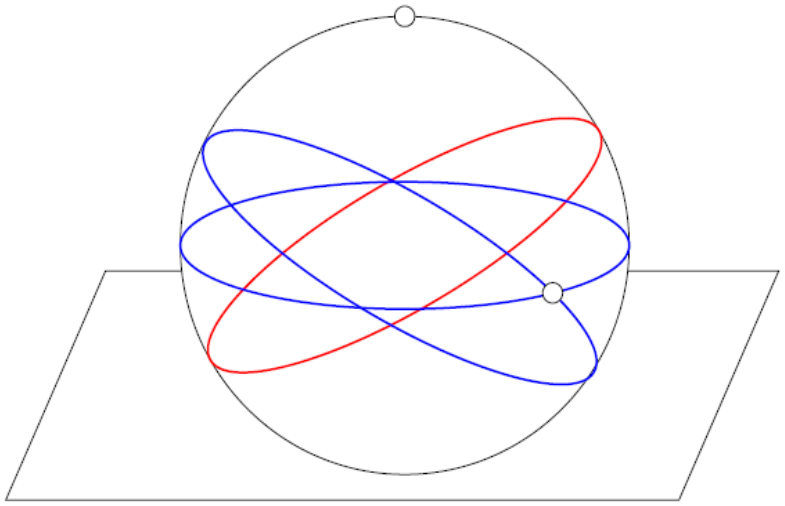
- Any two great circles intersect
  - Use great circles in place of vertical/horizontal lines

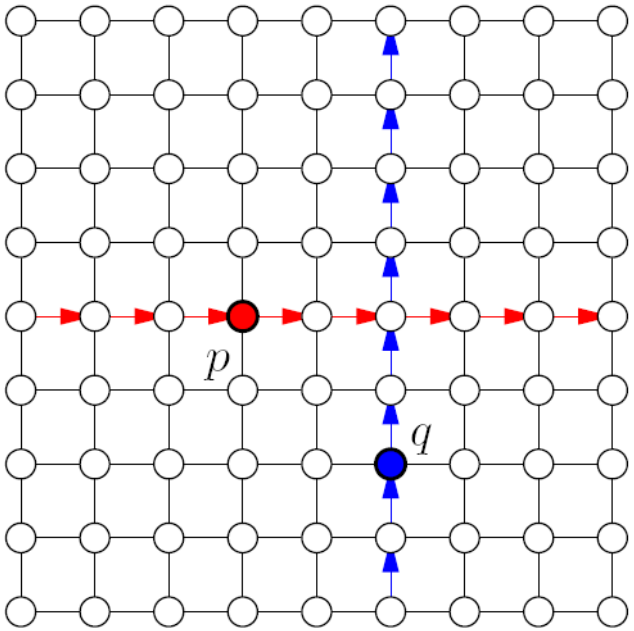




Spherical Double Rulings

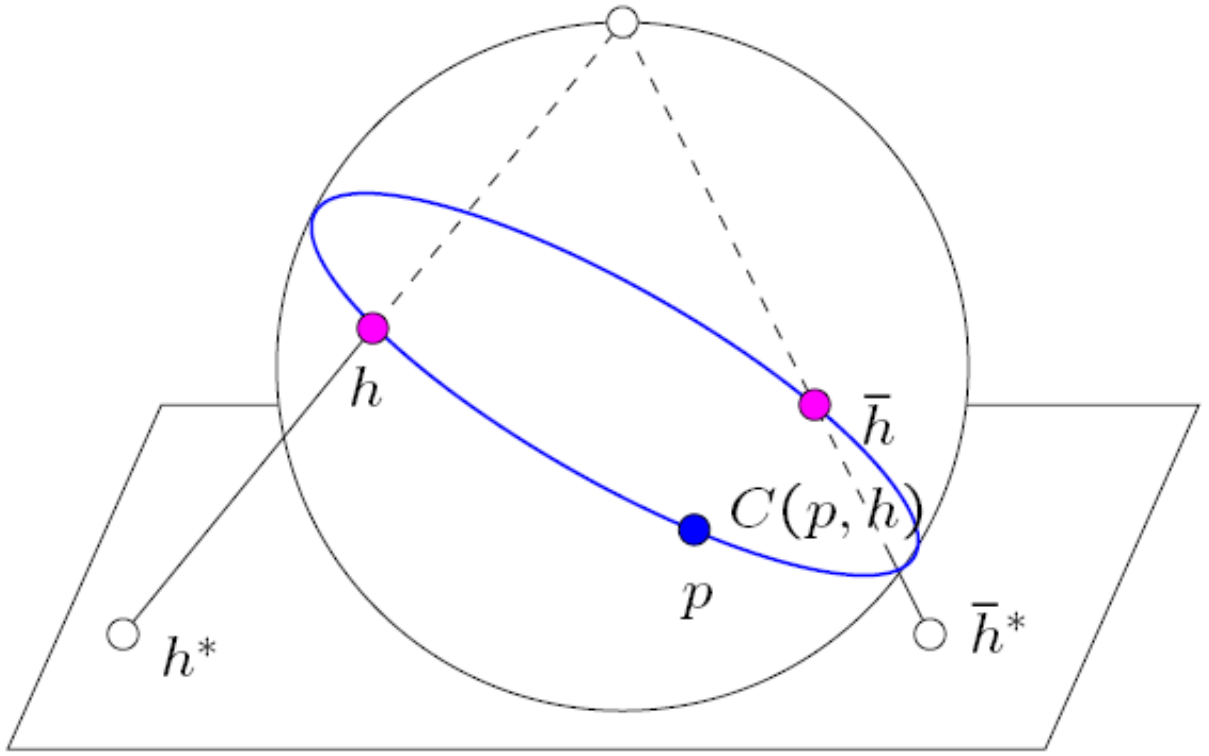
- One major difference with rectilinear double rulings:
  - Infinitely many great circles through a point
  - A lot more flexibility





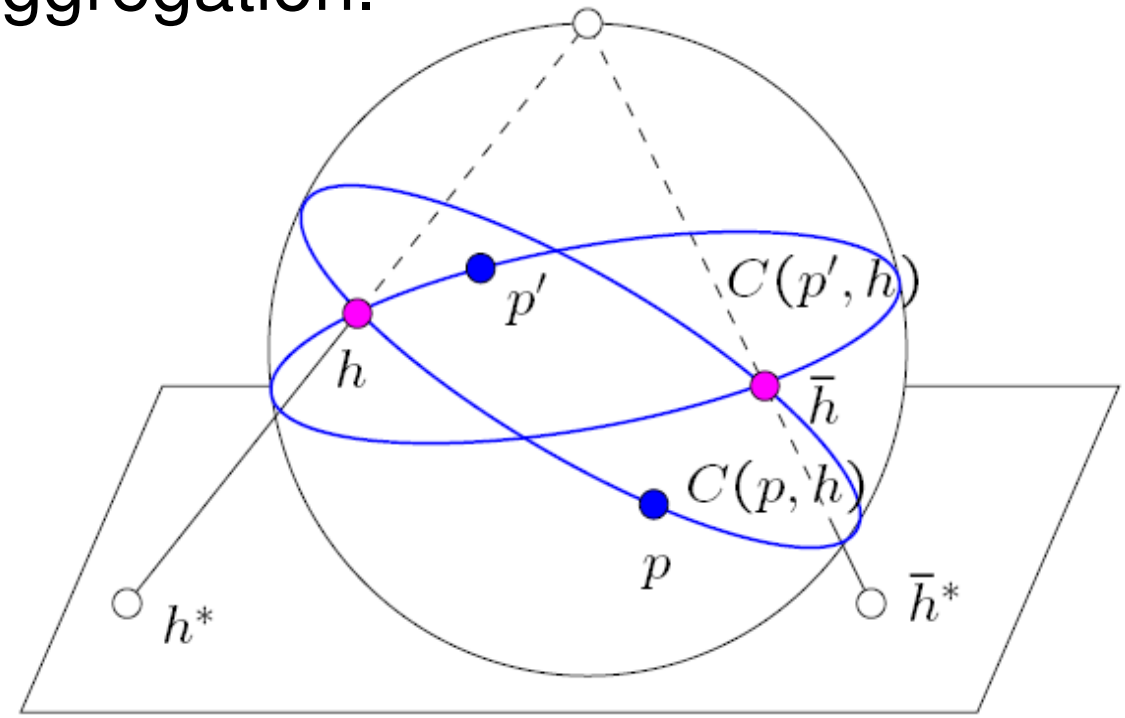
Data Replication

- Data centric hash function $h(T_i)=h_i$.
- Producer p replicates data along the great circle $C(p, h_i)$.

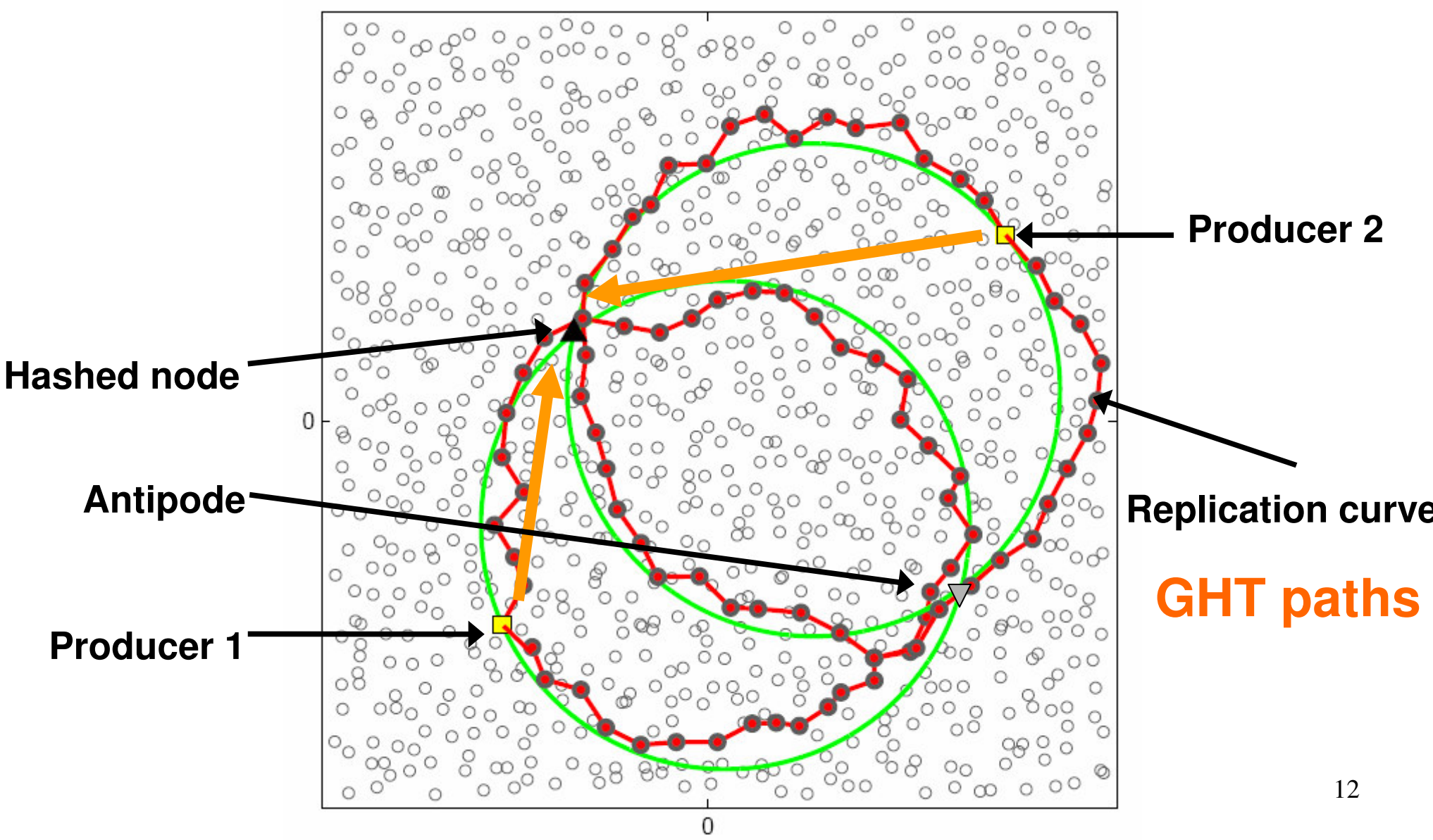


Data Replication

- Different producers with the same data type hash to different great circles, all passing through h, and its antipodal point $\bar{h}$.
  - Allow aggregation.



Replication Curve Examples



Data Retrieval

- Flexible retrieval rules
  1. GHT Style Retrieval
  2. Distance Sensitive Retrieval
  3. Aggregated Data Retrieval
  4. Full Power Data Retrieval

1. GHT Style Retrieval

- GHT still works
- Consumer q *wants data* T_i

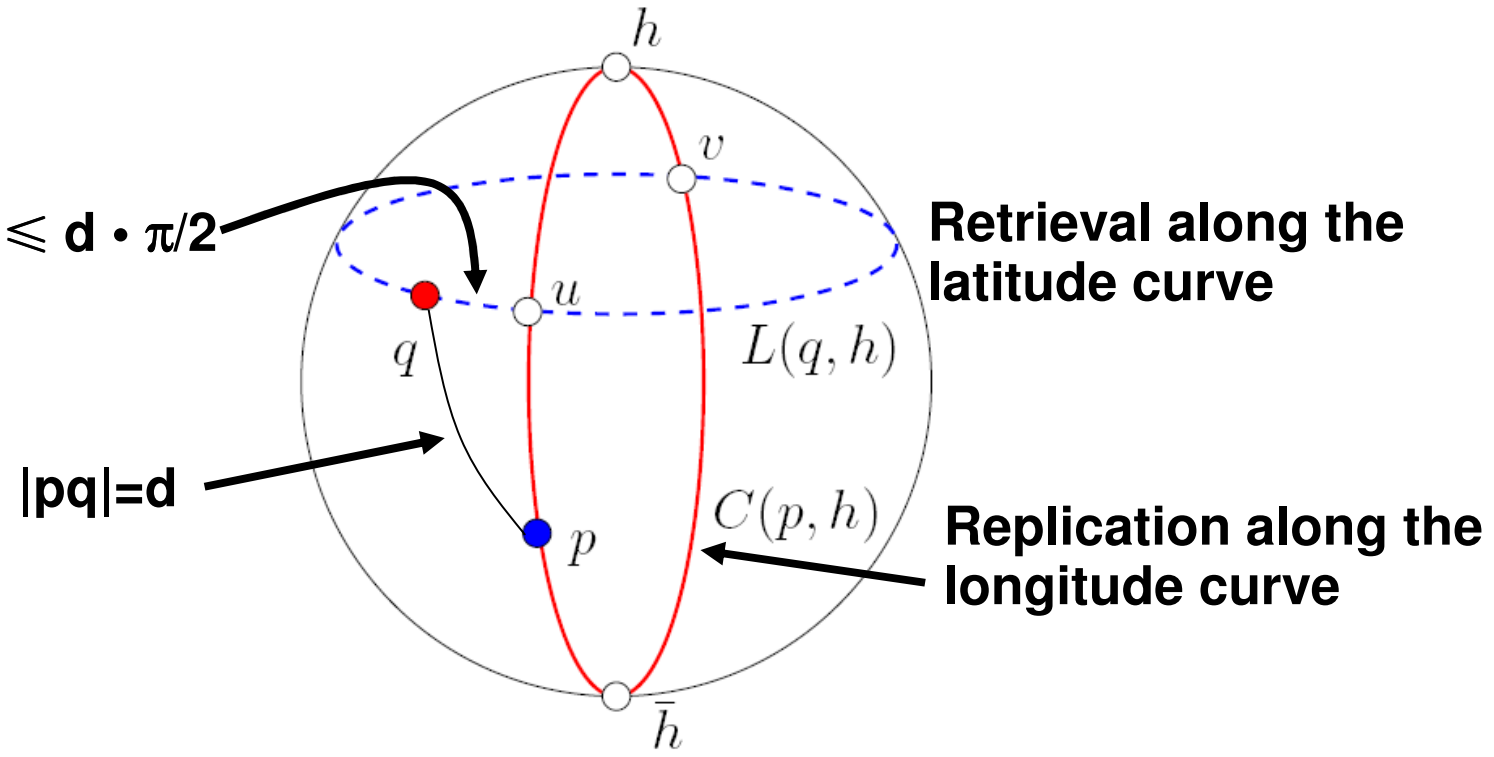
Consumer goes to hashed node h or its antipodal, whichever is closer.

2. Distance Sensitive Retrieval

- Distance Sensitive : If producer is at distance d from q, consumer should find data with cost $O(d)$.
  - Consumes less network resources
  - Users are likely to be more interested in immediate vicinity.
  - Lower delay --- Important in emergency response.

2. Distance Sensitive Retrieval

- Rotate the sphere so that hashed node is at the north pole.



If q is d away from p, the distance from q along latitude curve is ⩽ d • π/2.

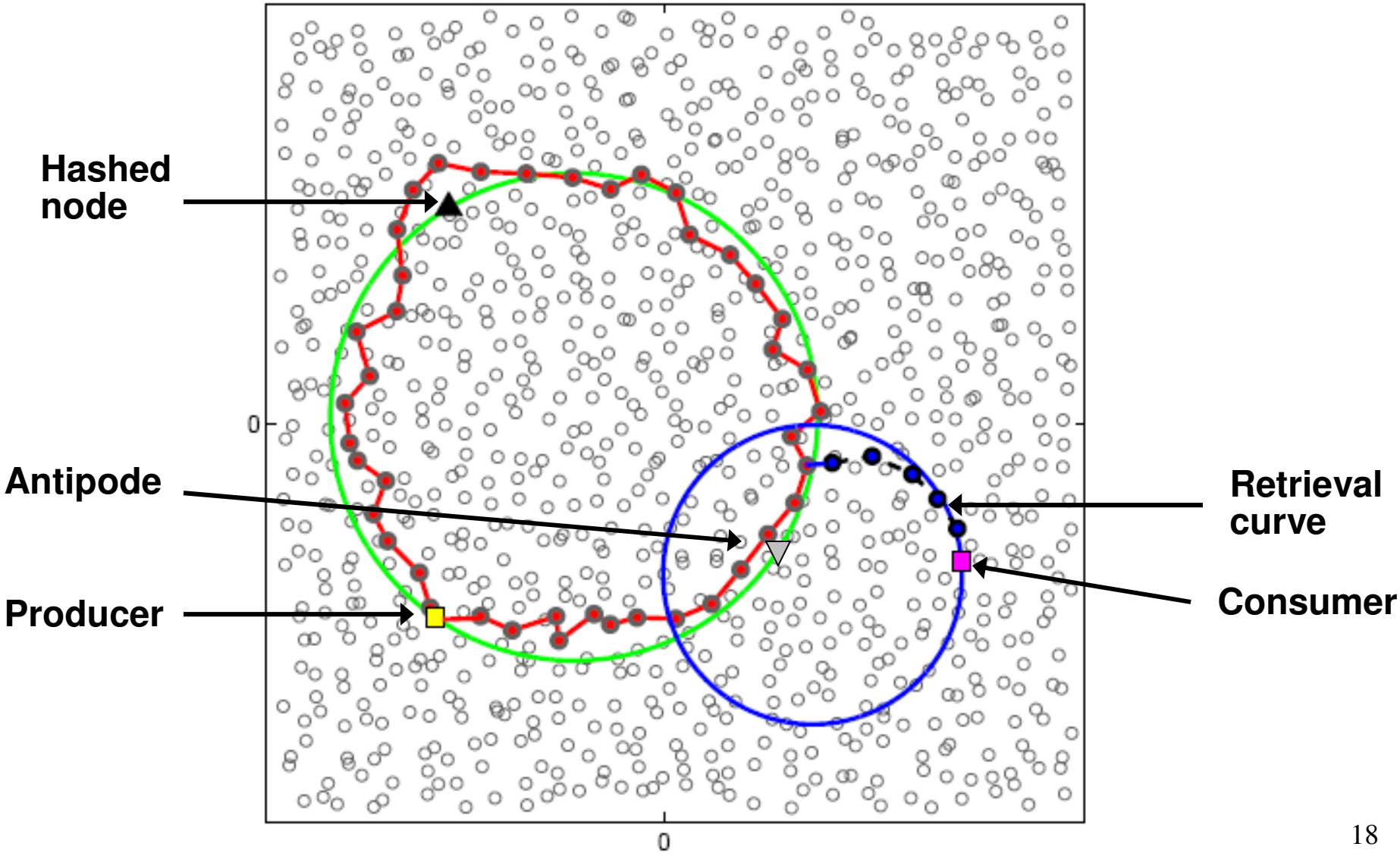
2. Distance Sensitive Retrieval

- Distance Sensitive : If producer is at distance d from q, consumer should find data with cost $O(d)$.

  Consumer q follows the circle with fixed distance to the hashed location.

- Wrong direction ?
  - Handled using a doubling technique
  - A random choice of direction works well in practice (we use this in simulations).

2. Distance Sensitive Retrieval



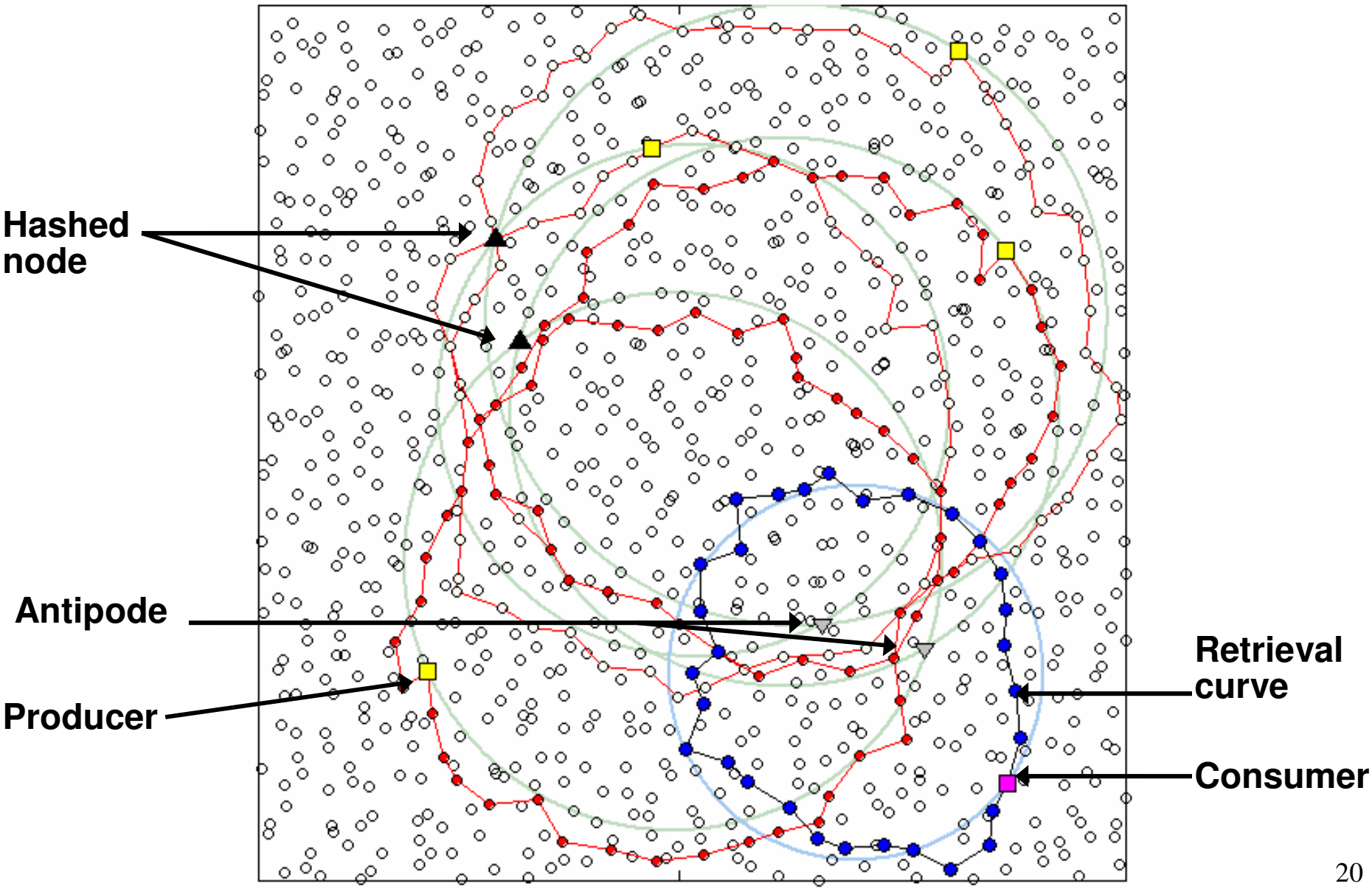
3. Aggregated Data Retrieval

- Consumer wants data of several Data Types $\{T_i\}$
  - E.g., monkey & elephant detections.

> Follow a closed curve that separates h_i and its antipodal point, for each data type T_i

  - Correctness: Any closed cycle that separates h_i from its antipodal intersects the producer curve.
  - Many such retrieval curves! ➔ more freedom for consumers and better load balancing.

3. Aggregate Data Retrieval



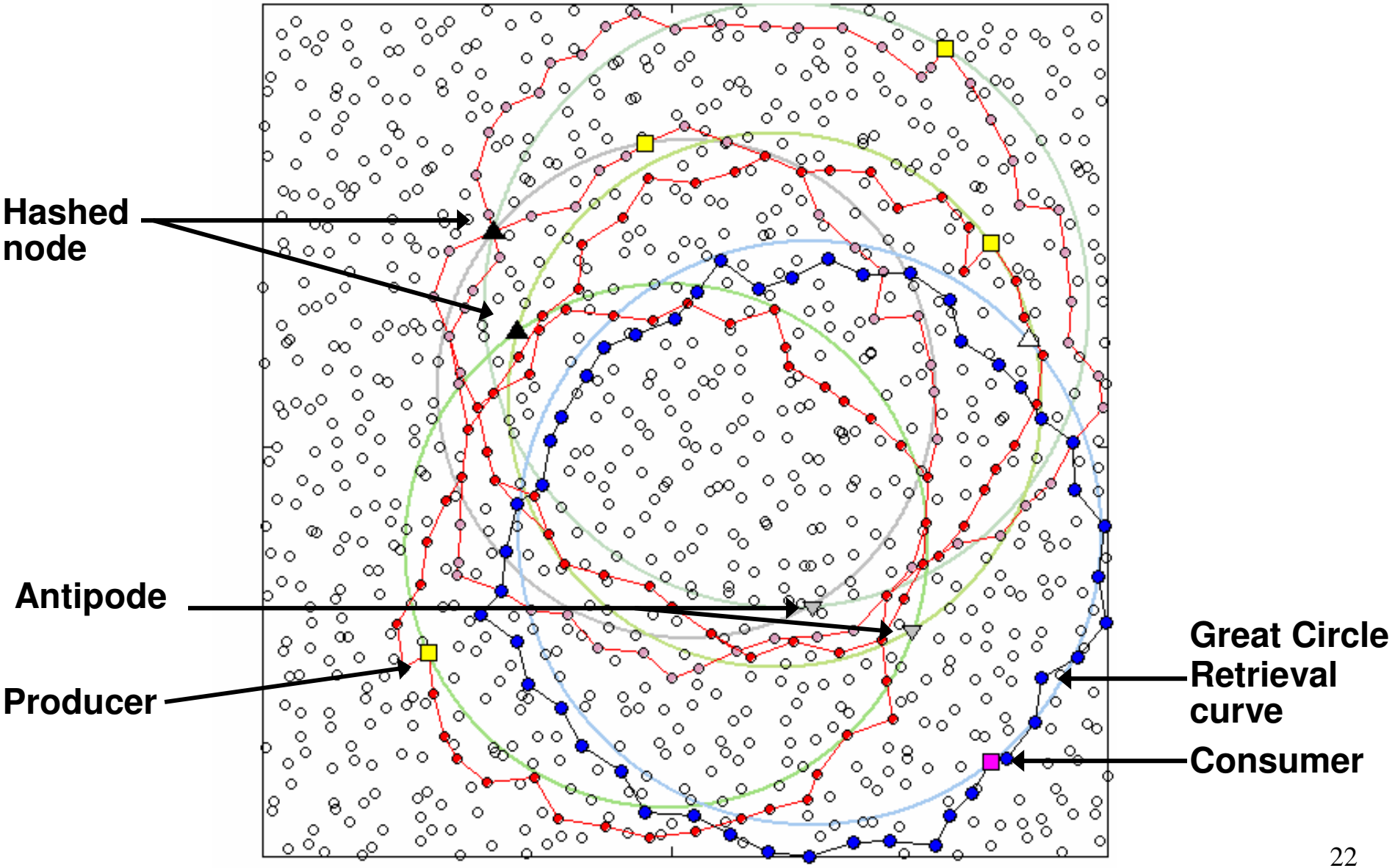
4. Full Power Data Retrieval

- Consumer wants all the data in the network

> Follow a great circle, retrieve all data.

  - Correctness : Any two great circles intersect
  - Many such great circles!

4. Full Power Data Retrieval

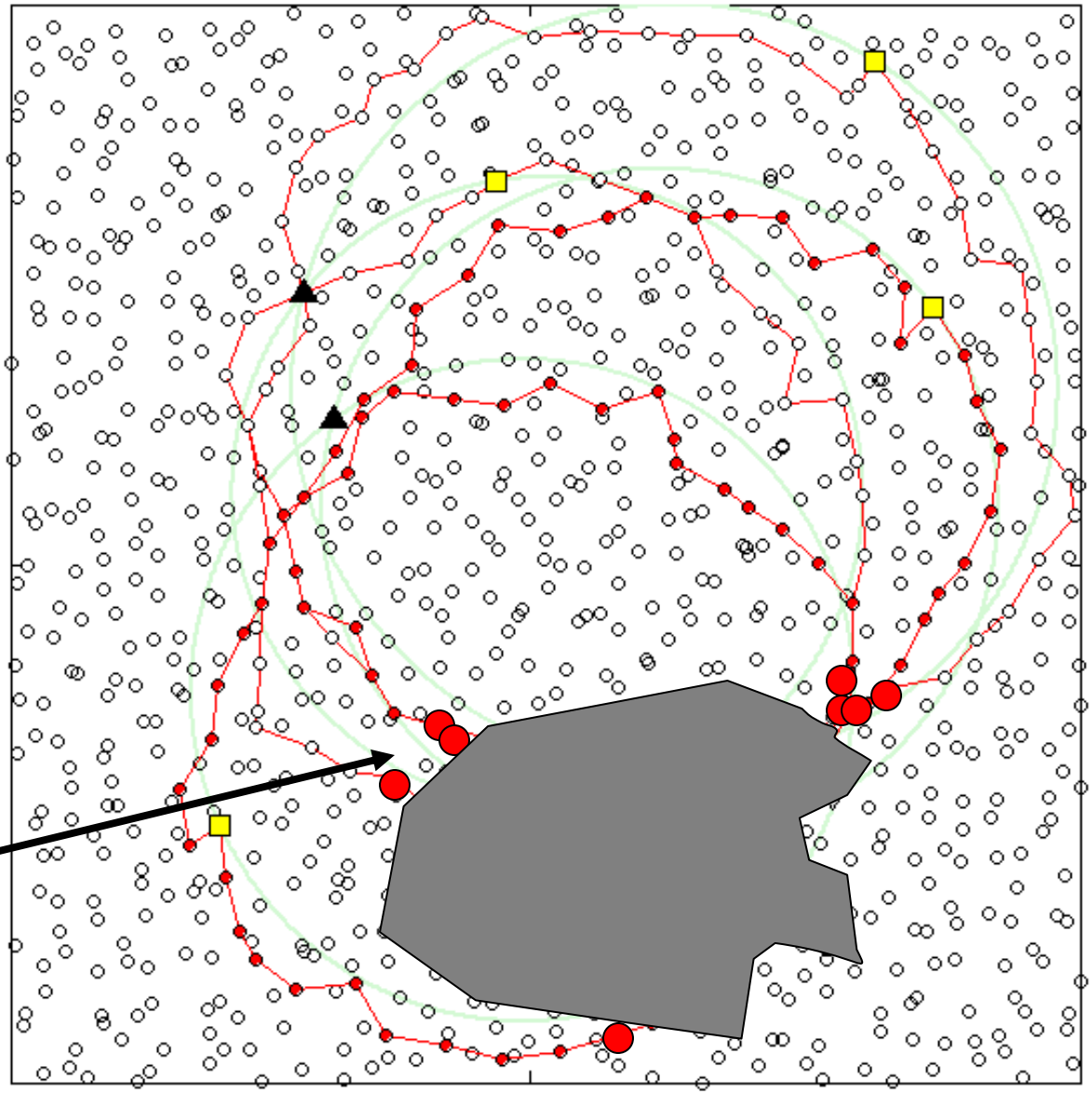


Local Data Recovery upon Node Failures

- When a group of nodes are destroyed,

> All the data on those nodes are available on the boundary of destroyed region.

Local Data Recovery upon Node Failures

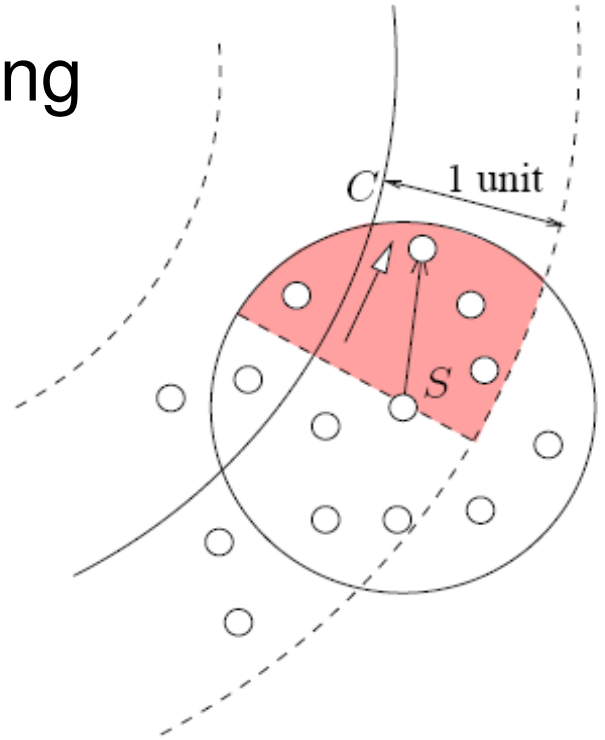


Survived Data Replicas on the boundary

Implementation

- How to forward data on a virtual curve ?
  - Use “Geographic Greedy forwarding on a Curve”

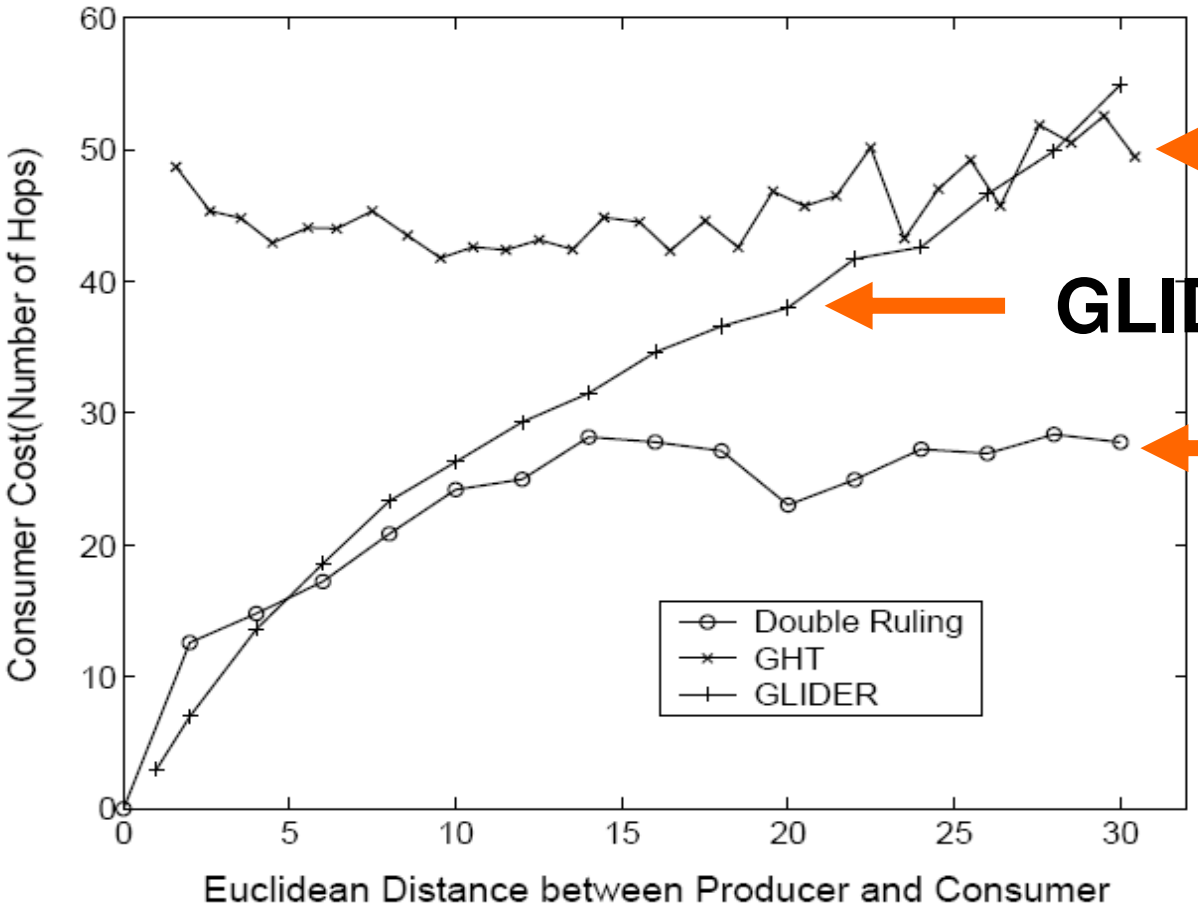
Badri Nath and D. Niculescu. Routing on a curve. *SIGCOMM Comput. Commun. Rev.*, 2003.



- The question of density
  - Is it always possible to forward ?
  - Simulation : A suitable 2-hop neighbor exists with high probability, for networks with avg degree $\geqslant$5.

Simulation: Distance Sensitivity

4200 nodes with average degree 8 per node.



GHT

GLIDER scheme

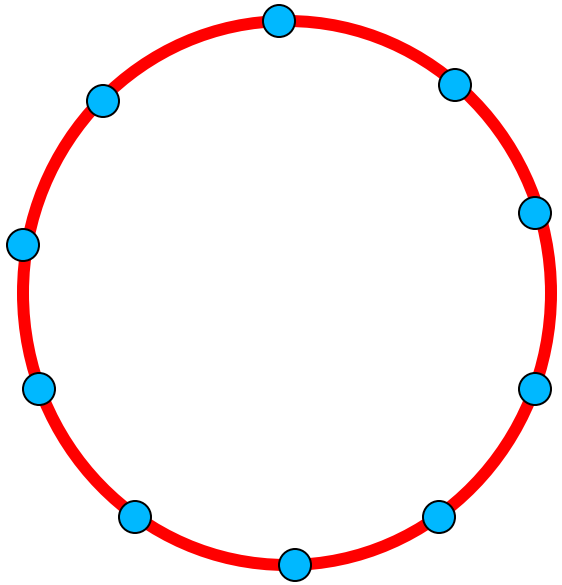
Spherical Double Ruling

Distance Sensitivity of queries

GLIDER based scheme : Q. Fang et. al. Landmark-based information storage and retrieval in sensor networks. *INFOCOM* 2006.

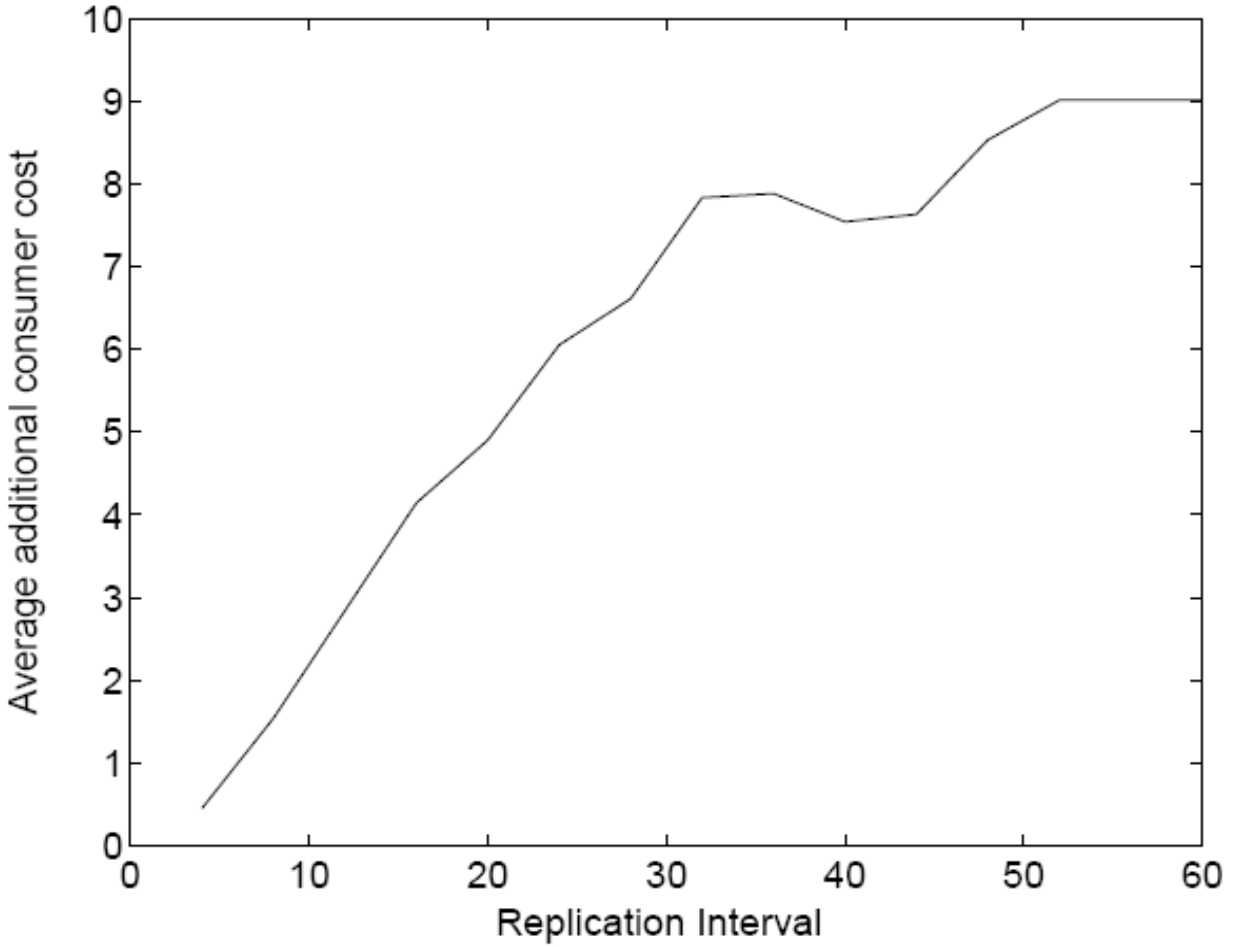
Simulation: Storage/Retrieval Tradeoff

Nodes on replication curve can store the data or a pointer to the actual data.



Replication Interval

Increasing Consumer cost

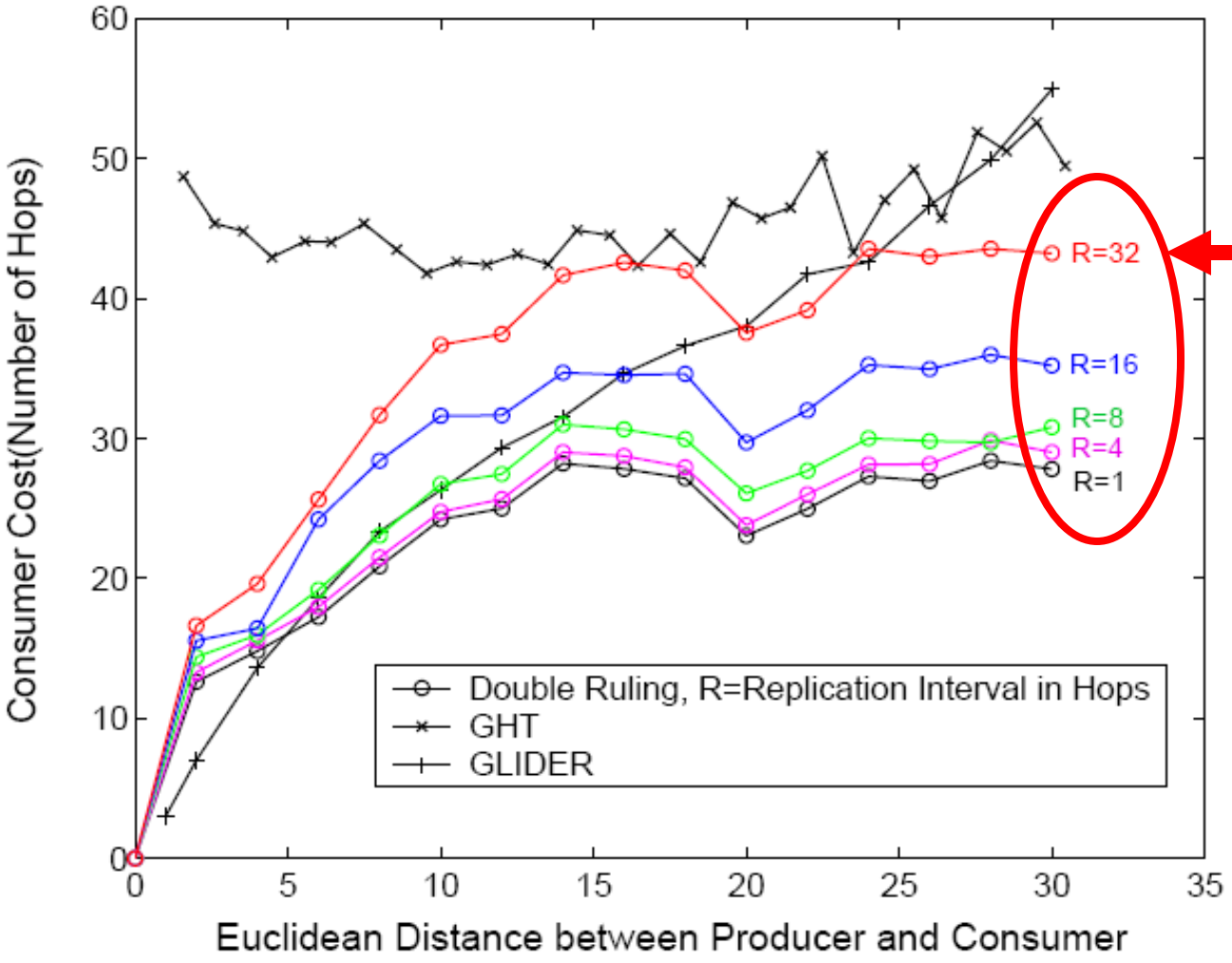


Larger Replication Interval

Decreasing Storage Cost

Simulation: Storage/Retrieval Tradeoff

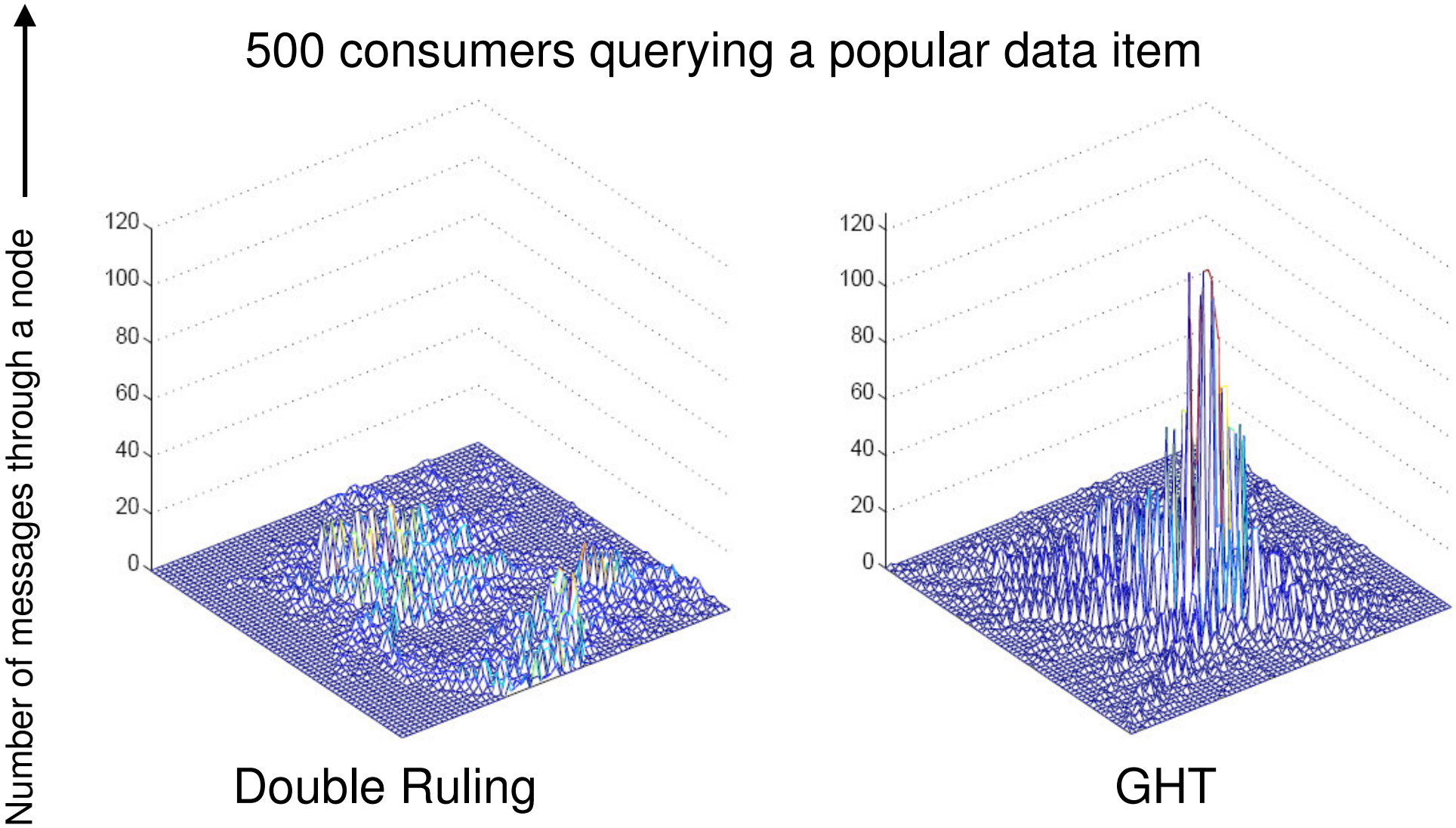
More storage, Lower retrieval cost.



Replication only on the hashed node and antipode.

Simulation: Load Balancing

500 consumers querying a popular data item



Discussion

- Data collection by mobile data mules.
  - Physically move along any retrieval curve.
- Advanced hashing schemes.
  - E.g., similar data types are placed nearby.
- Networks with holes.
  - Require special care.

Summary

- Flexibility in choice of retrieval method
  - Easier to hit a curve than a point.
  - many retrieval curves possible
- Better distance sensitivity
  - Even with sparse replication on producer path
- Better load balancing
- Better response to aggregate query:
  - Consistent performance with structured aggregate queries
  - GHT's cost increases linearly with number of data types
- Better average consumer costs

Questions

&

Comments..

References

S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R. Govindan, and S. Shenker. GHT: A geographic hash table for data-centric storage in sensornets. In Proc. *1st ACM Workshop on Wireless Sensor Networks and Applications*, 2002.

Q. Fang, J. Gao, and L. J. Guibas. Landmark-based information storage and retrieval in sensor networks. In The 25th Conference of the IEEE Communication Society *(INFOCOM'06)*, April 2006.

B. Nath and D. Niculescu. Routing on a curve. *SIGCOMM Comput. Commun. Rev.*, 2003.