Internet Measurement
& Analysis (IMA)

Topics in Internet Measurement

IP Geolocation

Prof. Georgios Smaragdakis, Ph.D.

slides courtesy of:

Chandrika Jayant

Ethan Katz-Bassett

Geolocation?

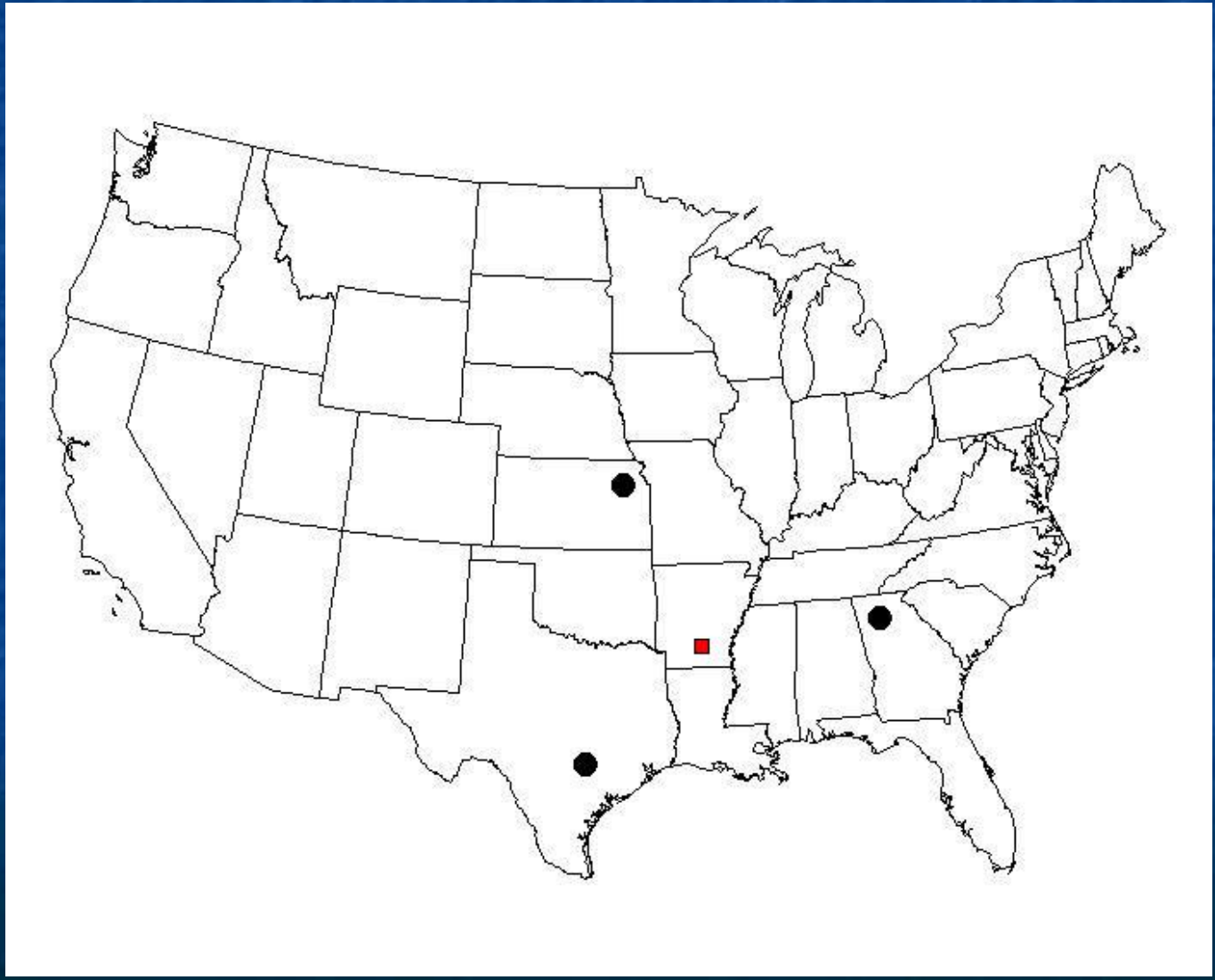
- Infer the geographic location of an Internet host
- Many applications would benefit from this information
- Advertising, location sensitive info
- Different levels of granularity

Constraint-Based Geolocation

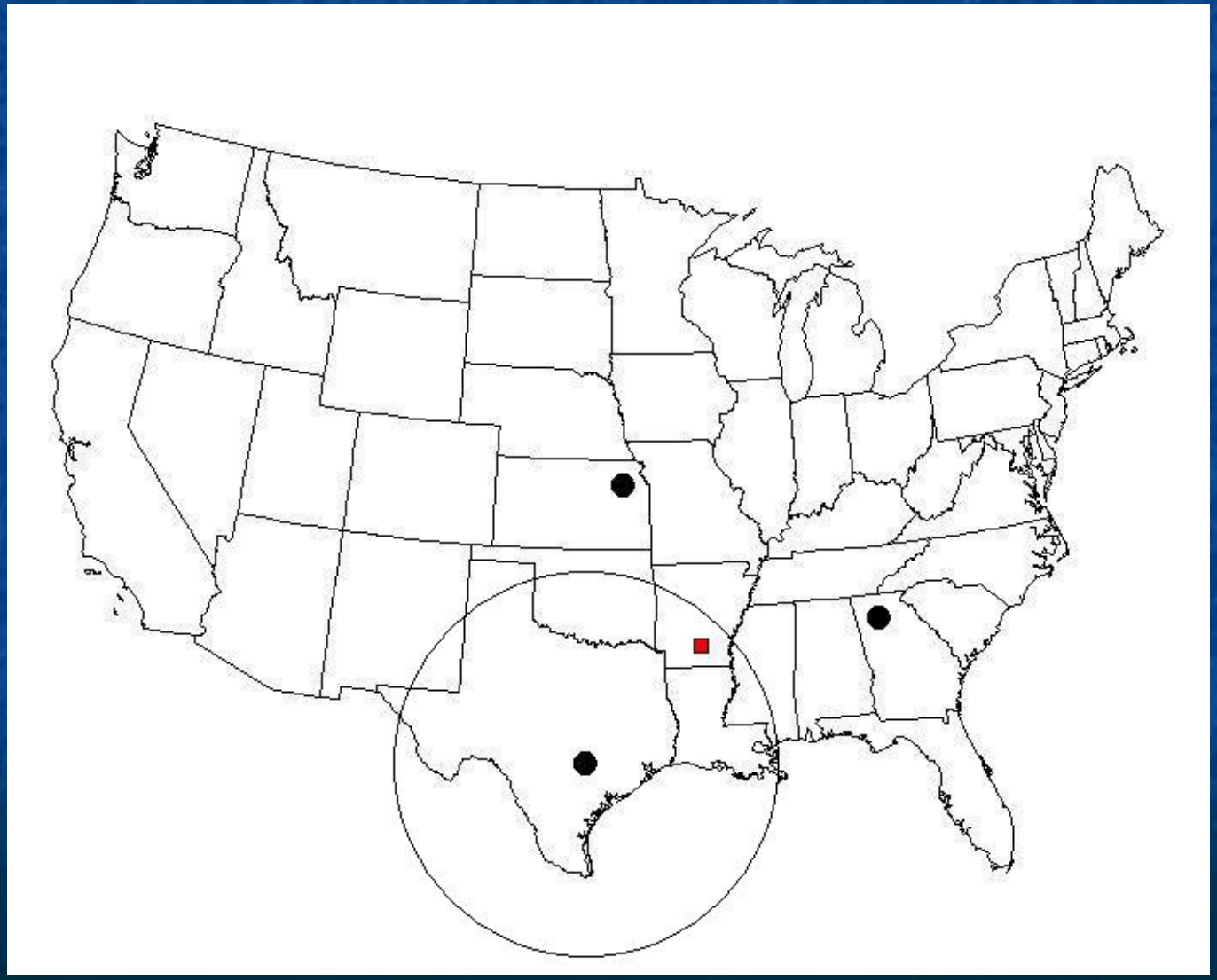
- Landmarks: Set of hosts with known locations
- Each landmark estimates distance to target
- Set performs multilateration using these distances

-Gueye, Ziviani, Crovella, Fdida (2004)

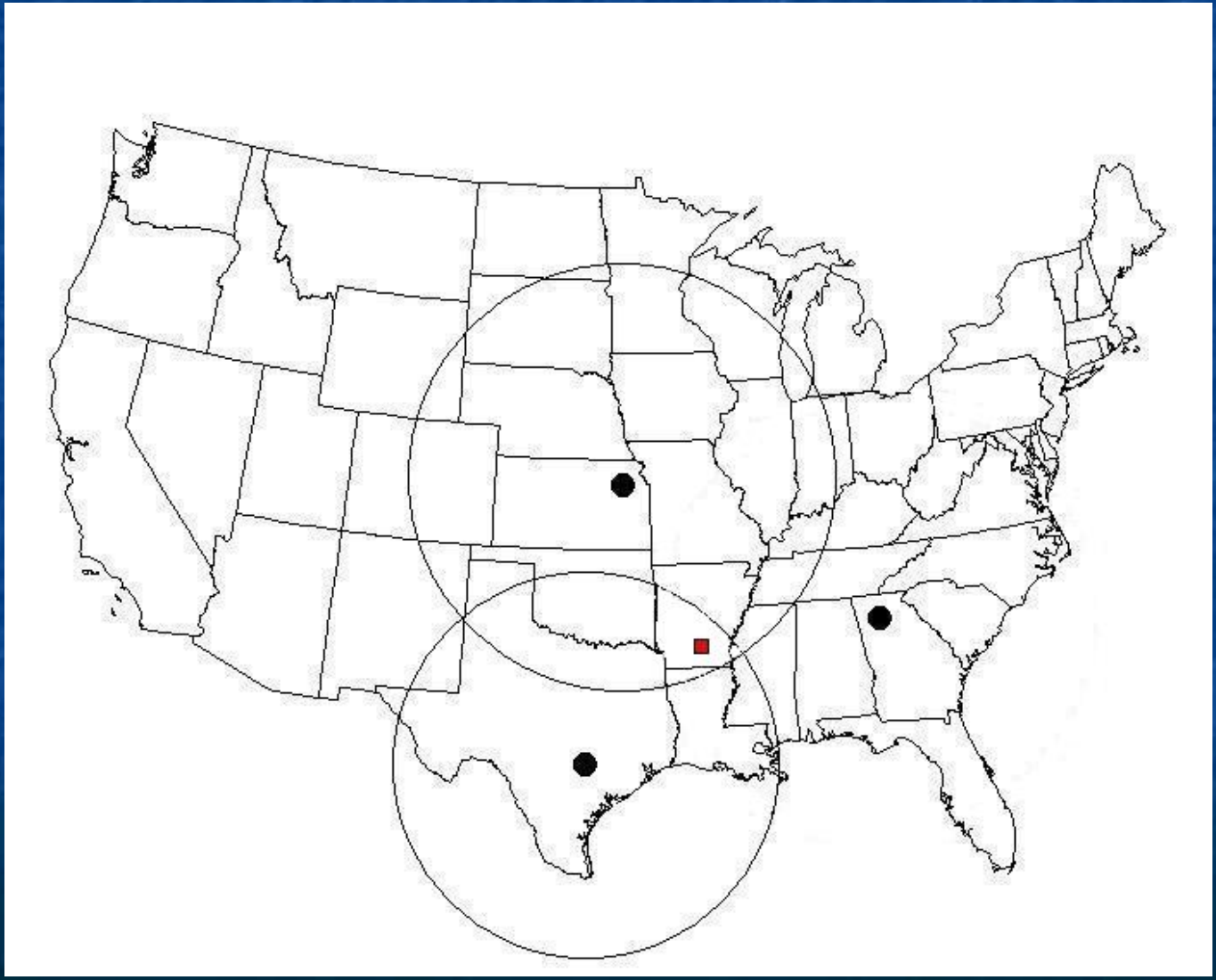
CBG Multilateration



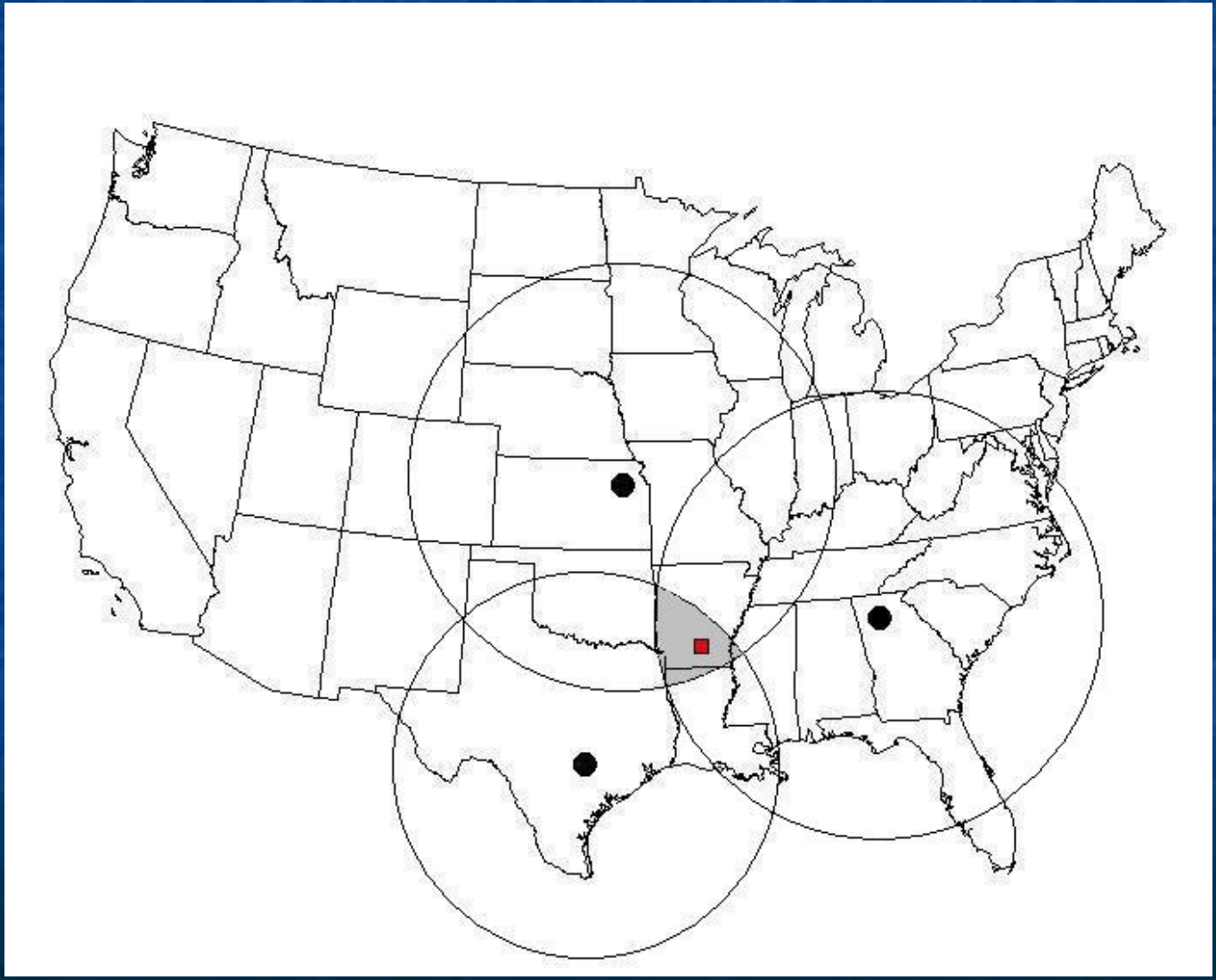
CBG Multilateration



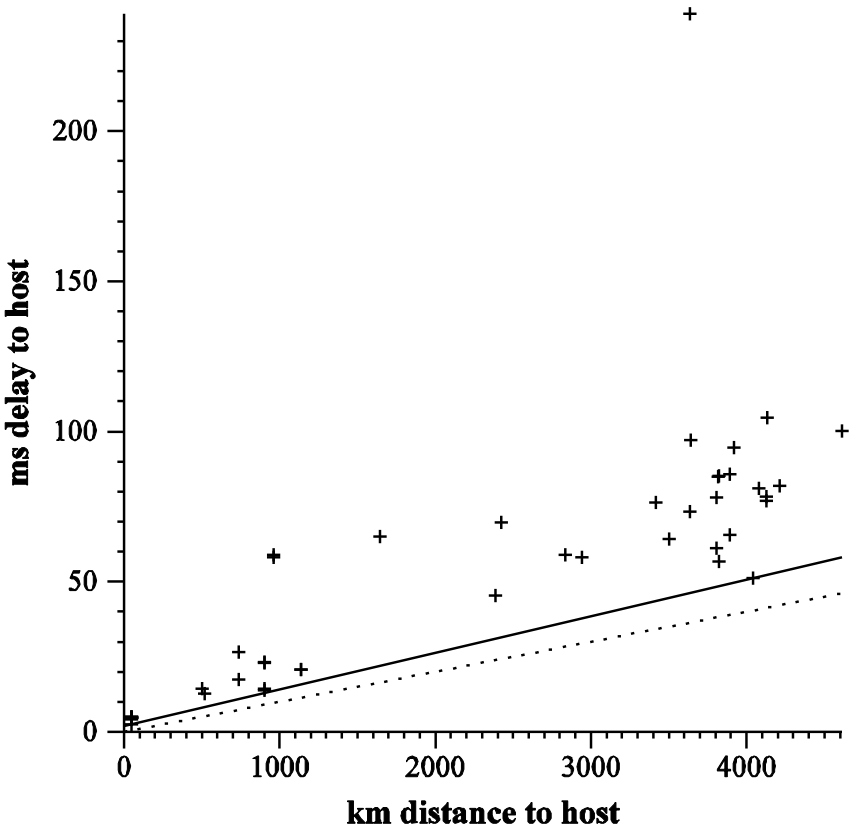
CBG Multilateration



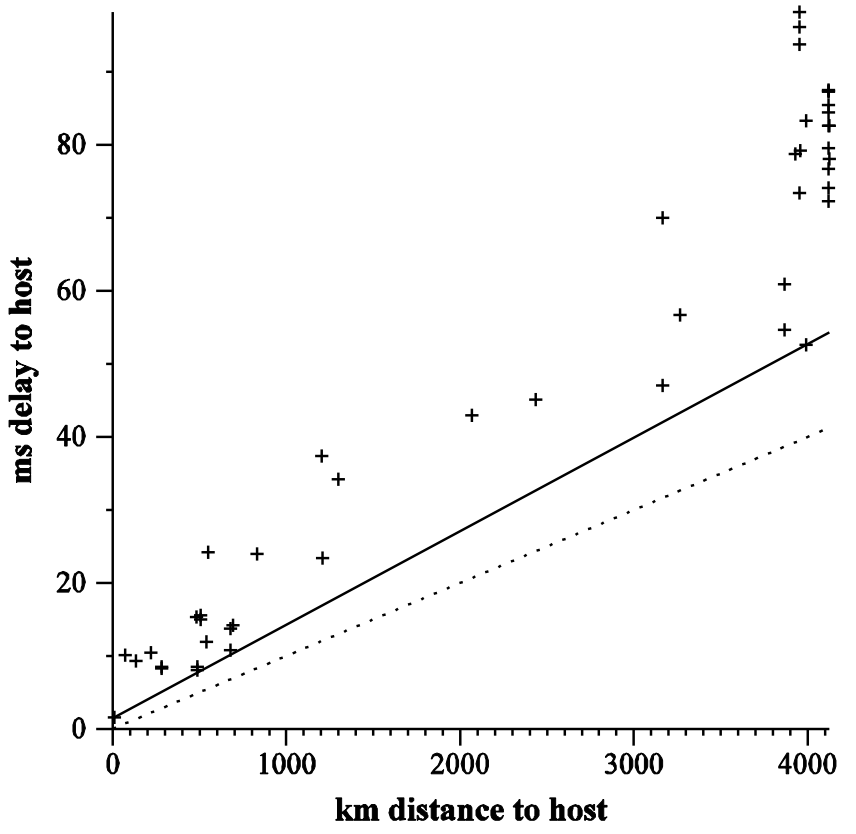
CBG Multilateration



CBG Bestline Distance Estimates



Bestline graph for pli2-pa-2.hpl.hp.com



Bestline graph for planet1.scs.cs.nyu.edu

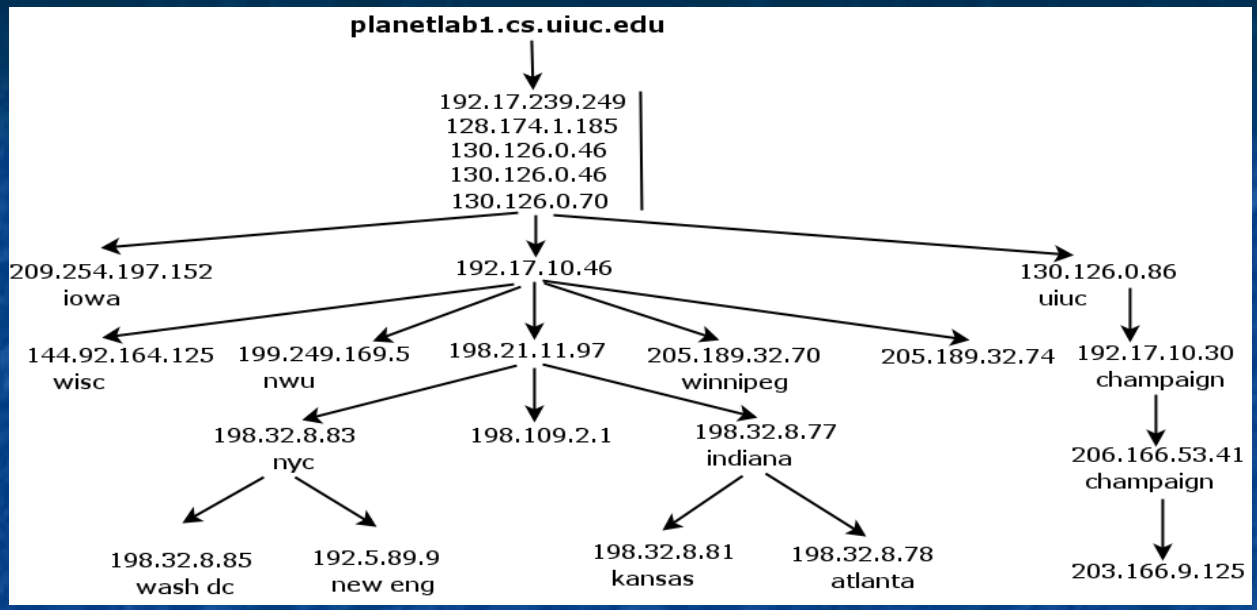
CBG Breakdowns (CBGB's)

- Estimates are not tight and vary widely → large confidence regions, need many probes to get a few tight ones
- No better at estimating training set vs. other hosts (in general)
- More data trained on, worse accuracy (in general)
- Still underestimate some distances

Our Approach

- Intuition: Targets that have similar routes have similar delay → distance conversions
- Use route info to achieve more accurate estimates
- Want to fit into CBG framework
- 2 main techniques, still using bestline fit:
  - Path-Based
  - Router-Based

Path-Based Estimation



- Landmark learns routes to its training set
- Traceroute target up to TTL = x
- Find longest partial path shared with a subset of training hosts
- Calculate bestline using only this subset

Router-Based Estimation

- Landmark learns routes to its training set
- Send packet to target with TTL = x
- Find subset of training hosts with paths through this router (do alias resolution)
- Calculate bestline using only this subset
- In practice, use $x_1, x_2, \ldots, x_n$

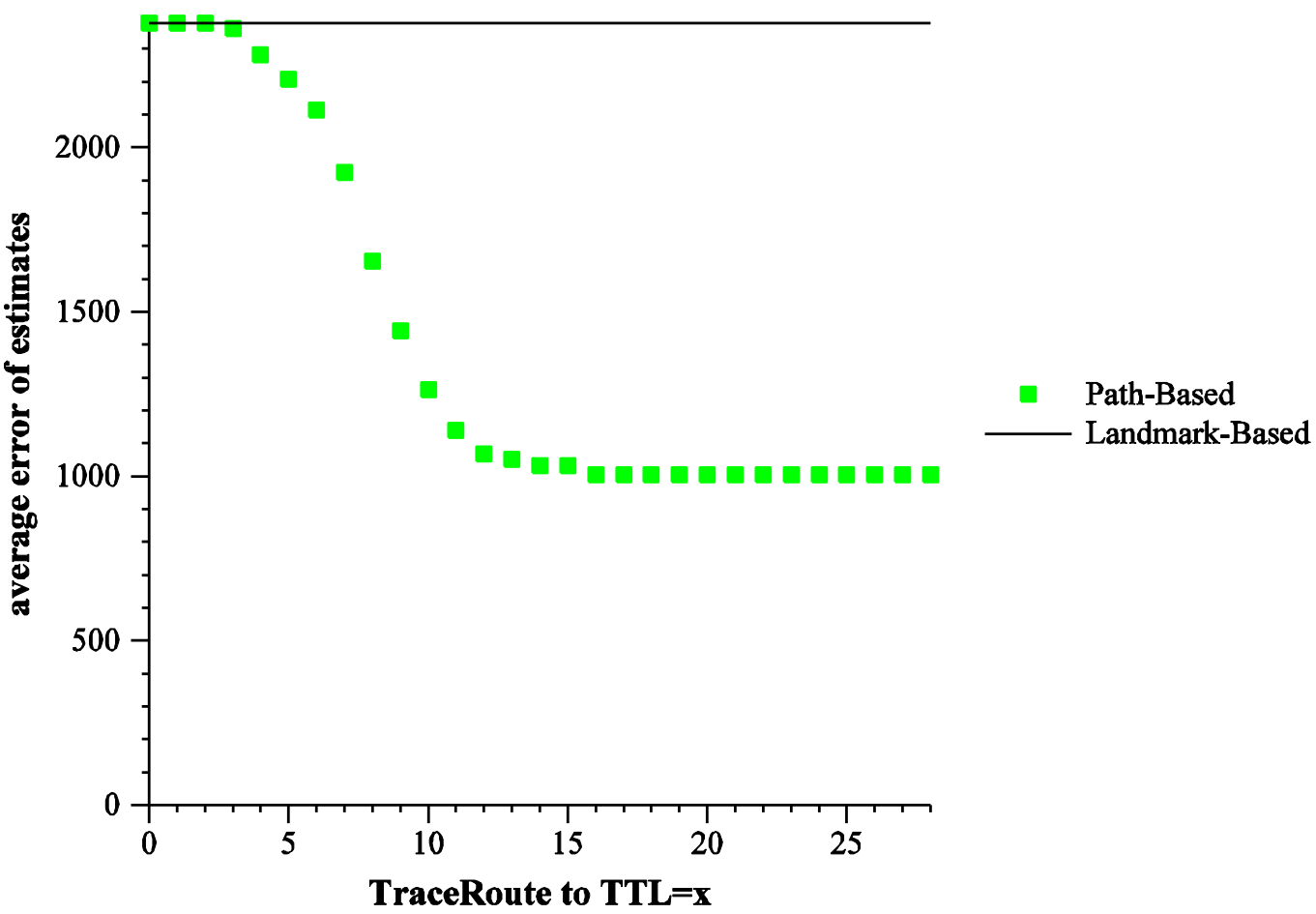
PlanetLab Experiments

- 110 PlanetLab hosts in North America
- Lat/long available for each
- Used Scriptroute to gather delay and routes between hosts
- 26 landmarks (after munging)
- Path-Based used TTL up to 12
- Router-Based used TTLs (12,9,6)

Map of Landmark Hosts (26)

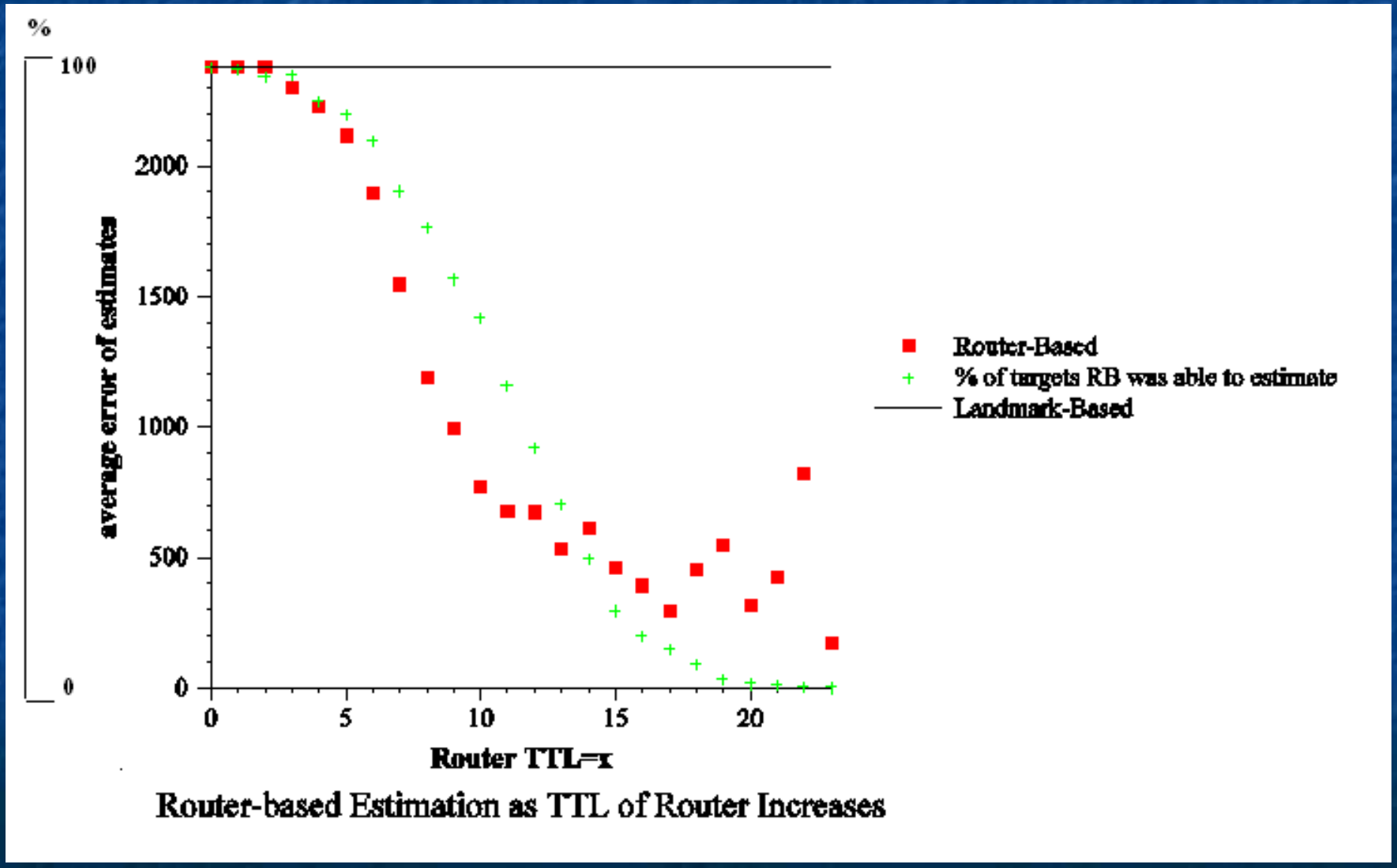
map_landmarks

Path Length vs. Accuracy



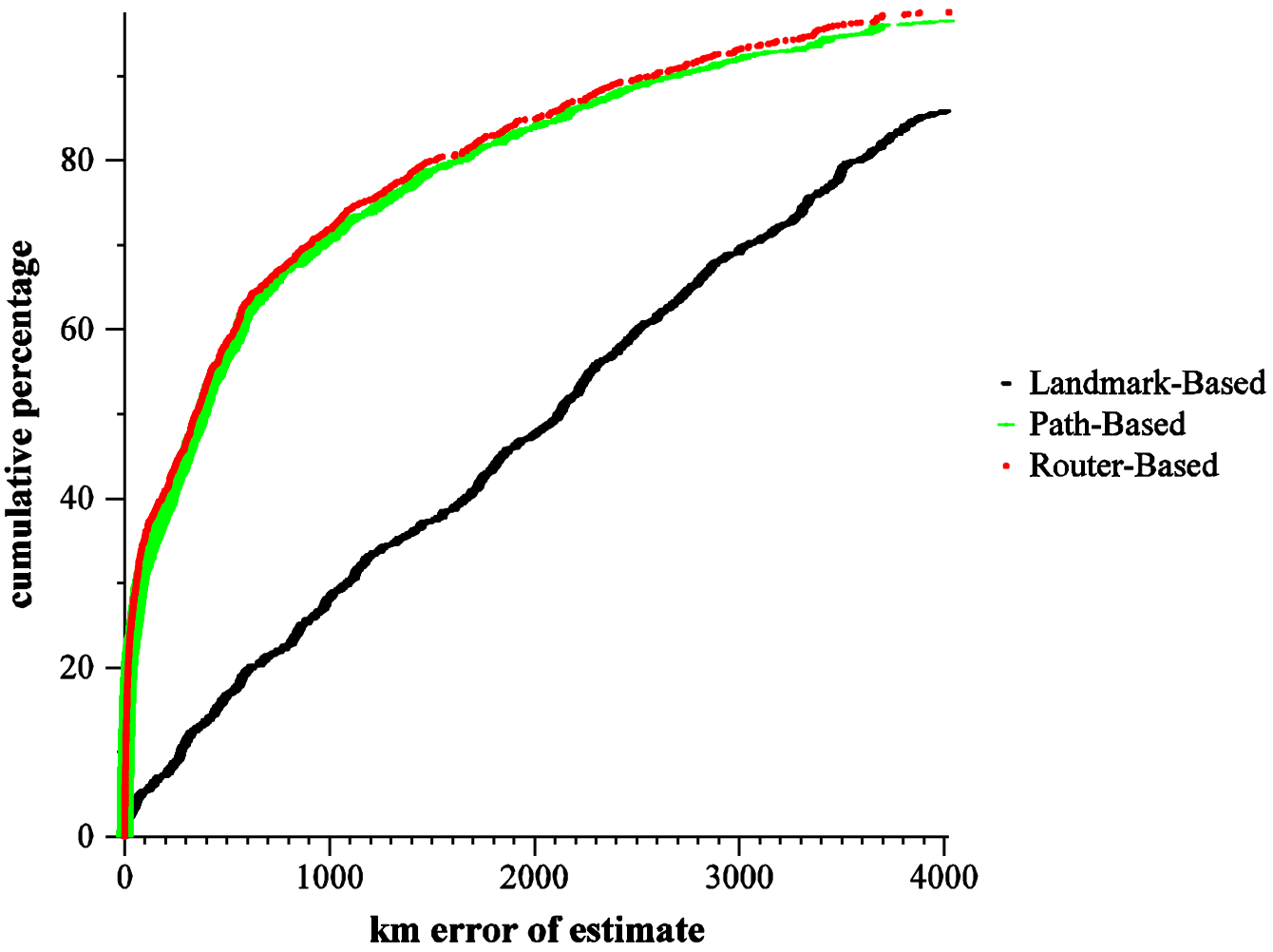
Path-based Estimation as # of Hops Measured Increases

Router TTL vs. Accuracy

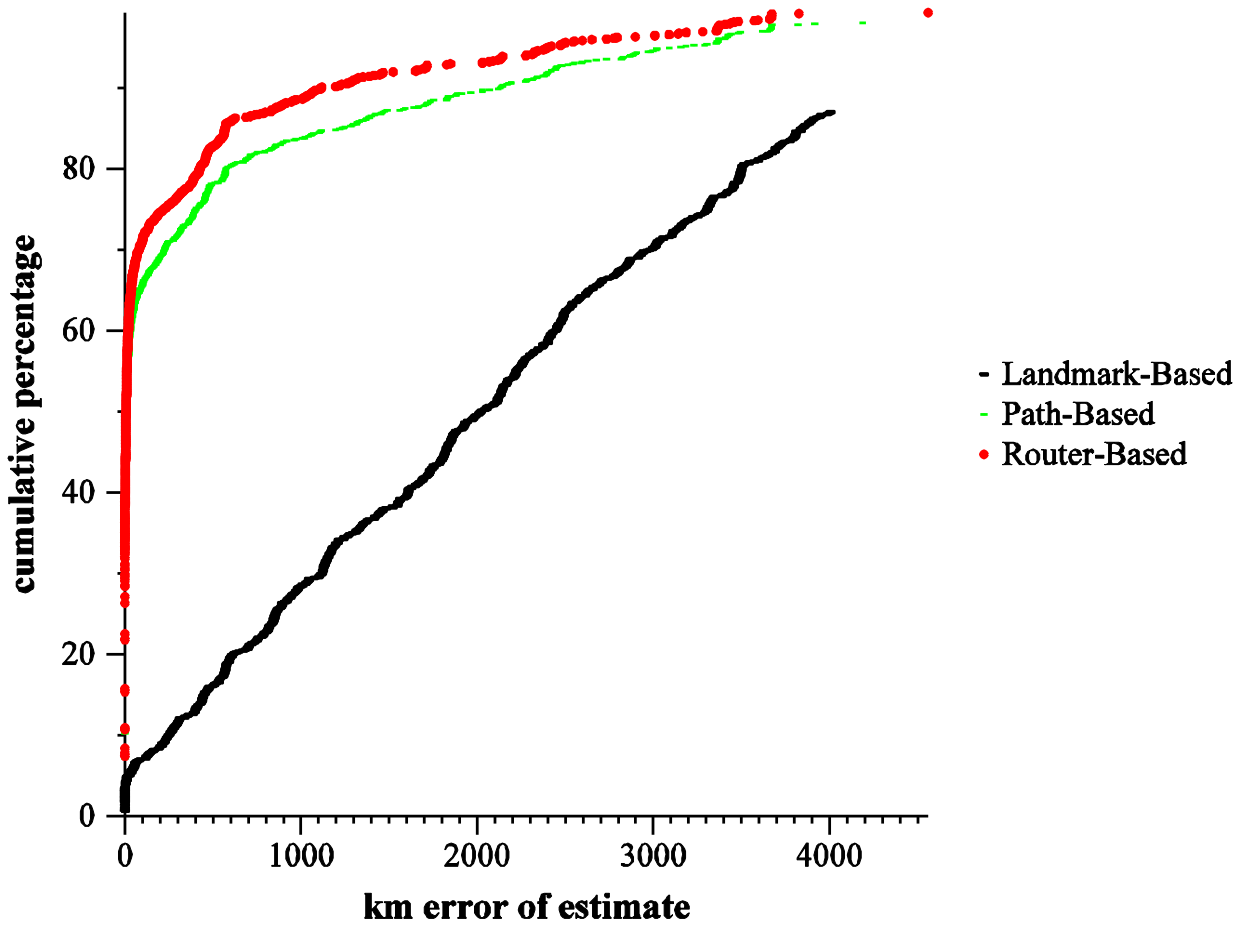


Router-based Estimation as TTL of Router Increases

Overall Accuracy of Estimations



Estimations to Collocated Targets



Estimations to Hosts Collocated with Landmarks

Conclusions

- Geolocation has powerful potential
- CBG is interesting, but needs improvement
- Route information improves accuracy of distance estimates
- With more accurate estimations, likely need fewer landmarks to locate targets

Reading

"Towards IP Geolocation Using Delay and Topology Measurements", E. Katz-Bassett, J. John, A. Krishnamurthy, D. Watherall, T. Anderson, and Y. Chawathe. IMC 2006