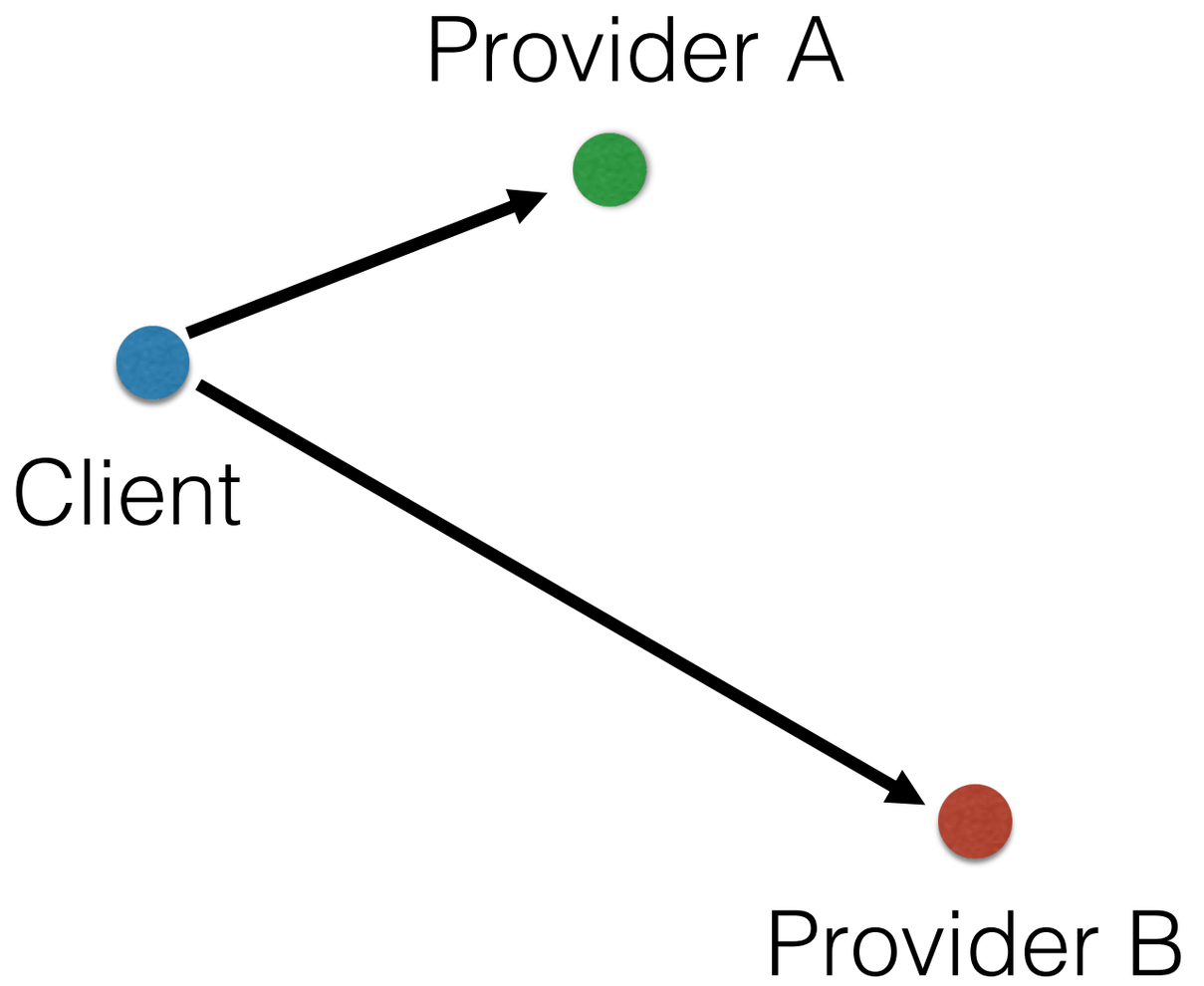
Fury Route

Leveraging CDNs to Remotely Measure Network Distance

Marcel Flores
Alexander Wenzel
Kevin Chen
Aleksandar Kuzmanovic

Relative network distance

- Which provider is closer to the client?
- Enable systems to make a best-guess prior to the first communication.
- Can we do it without infrastructure or host participation?



Relative network distance

- Peer and server selection.
- Large scale Internet measurement, allows understanding of hosts beyond control.
- Clients which can't be subjected to direct measurements (*i.e.* traffic sensitive, unreachable).

Use CDN replica selection

- CDNs provide end users with a "nearest" replica.
  - Helps to reduce user-experienced latency.
  - Often selected via DNS.
- Providers include: Google, Edgecast, Alibaba, CDN77, CloudFront, CDNetworks, Adnxs
  - Varying response granularity

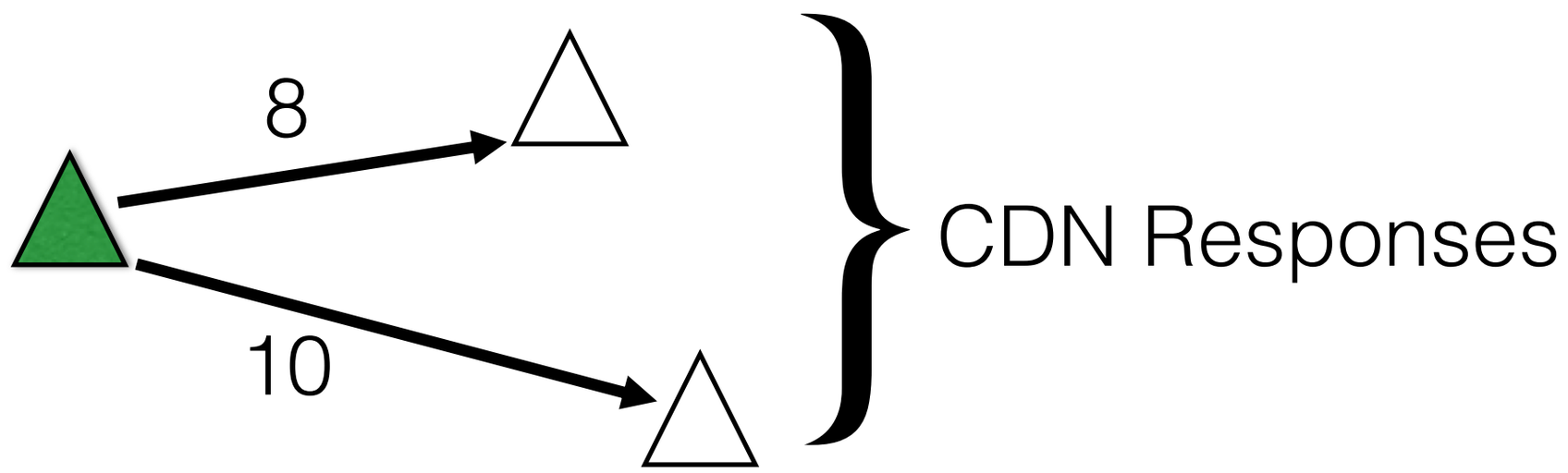
EDNS Client Subnet

- Allows DNS requests to include an origin subnet.
  - Simplifies replica selection procedure.
- Further provides a *scope,* indicating the specificity of the response.

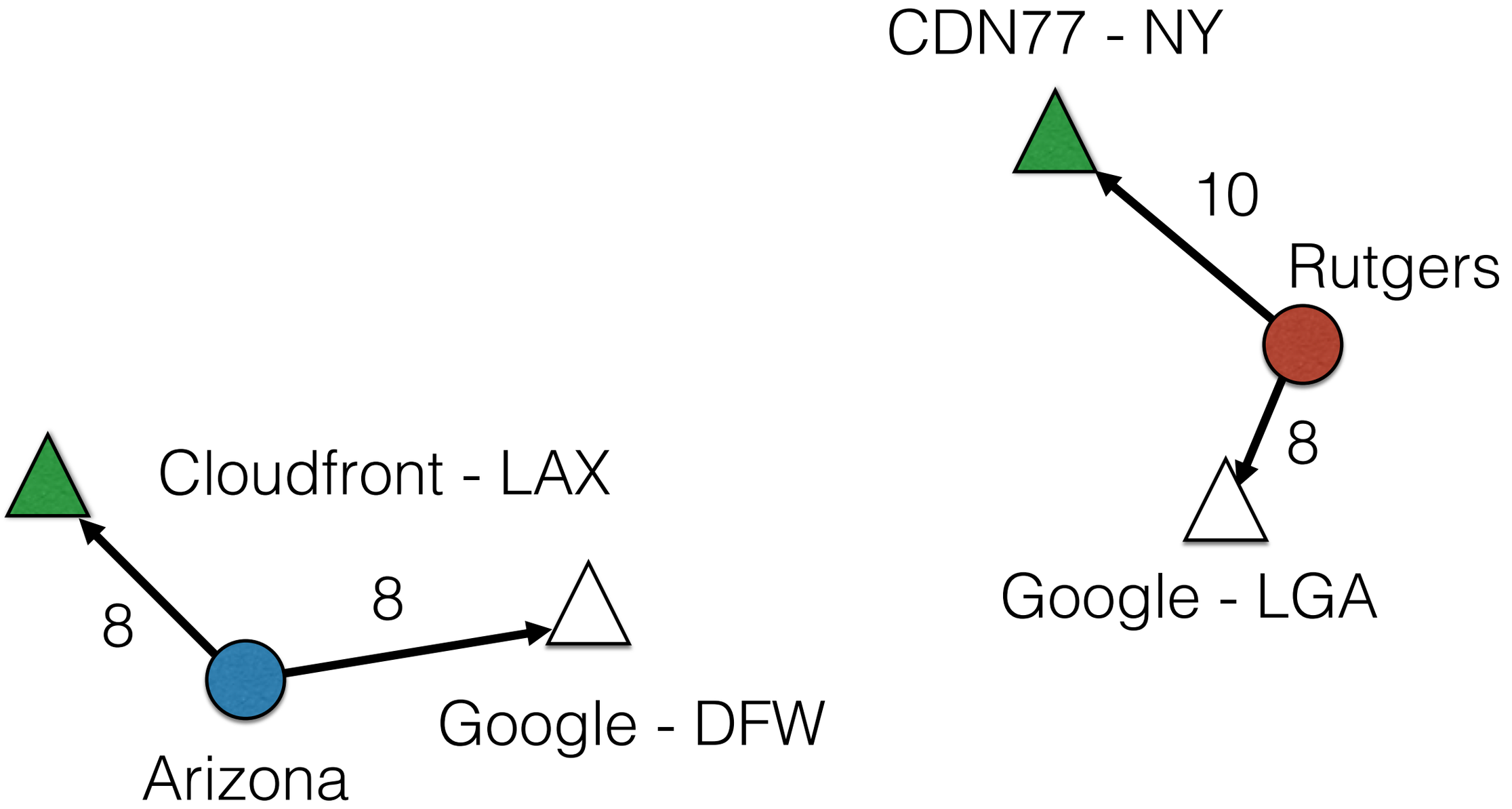
Allows for queries as arbitrary hosts.

Build a graph

- Construct a directed graph G = (V, E)
  - V = Set of hosts (client, provider, replicas)
  - E = Set of of CDN replica response relationships
- weight(A, B) = 32 - *scope* (a, b)



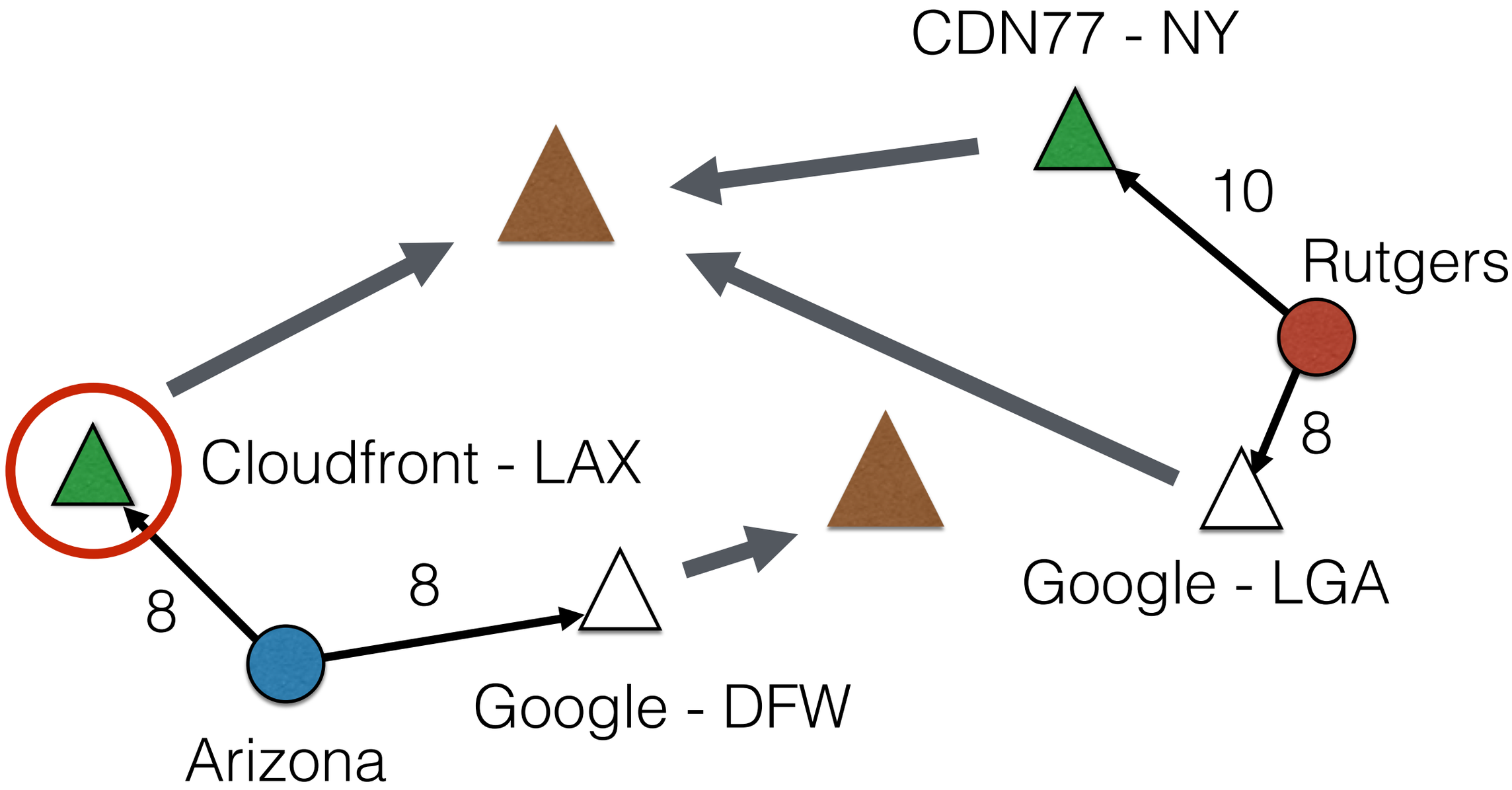
Build a graph



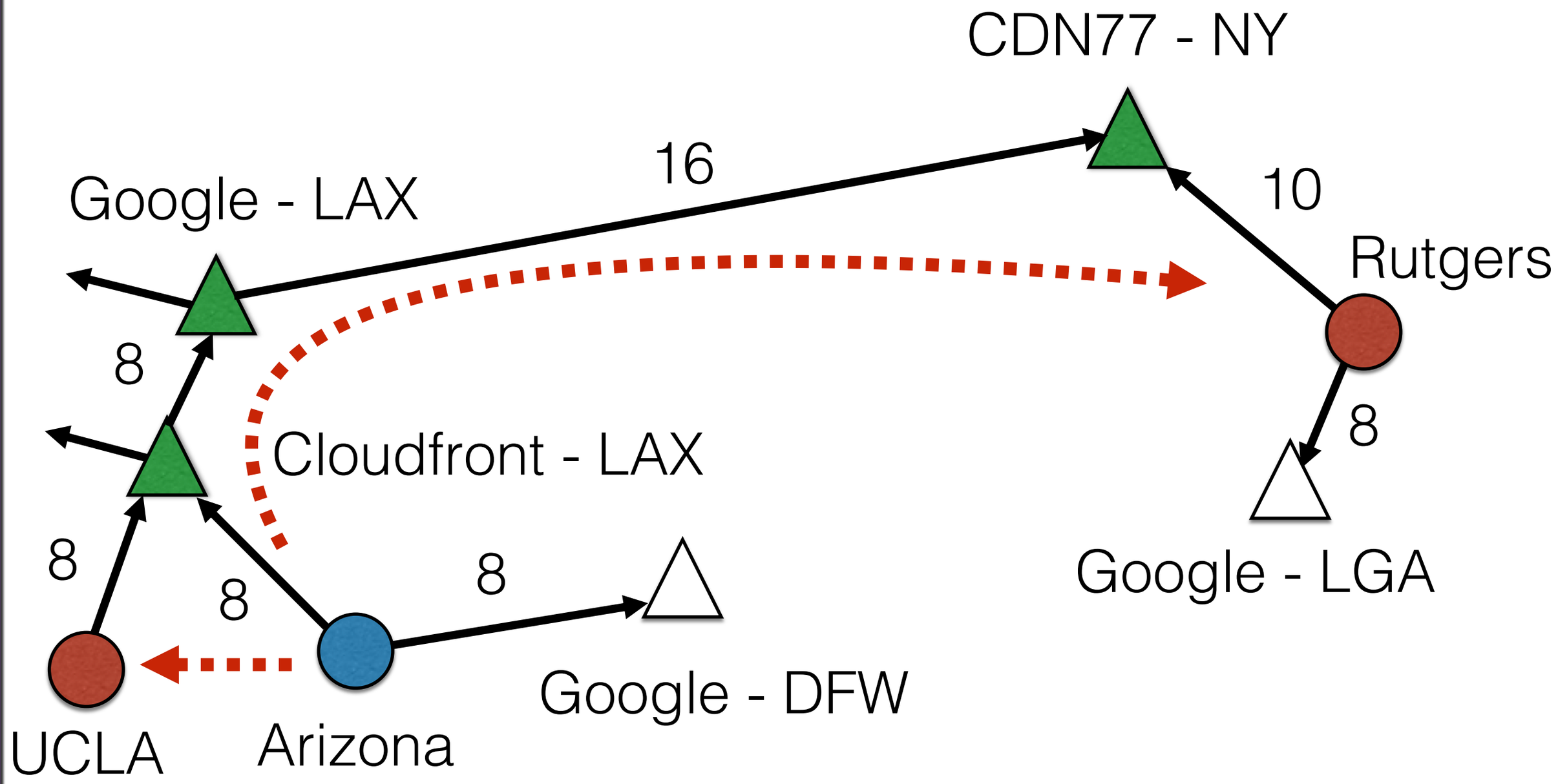
Forward progress

- How do we know which candidate to select?
  - Ask coarse-grained providers which replica they would serve to each candidate.
    - Further ask which replicas they would serve the target set.
- Measure the overlap in responses.

Build a graph



Build a chain



Why does it work?

- Ultimately reveals the underlying infrastructure of the CDN Deployments.
  - Low-density deployments will indicate large distances.
  - Not all deployments are the same.
- ECS responses are more than just noise.

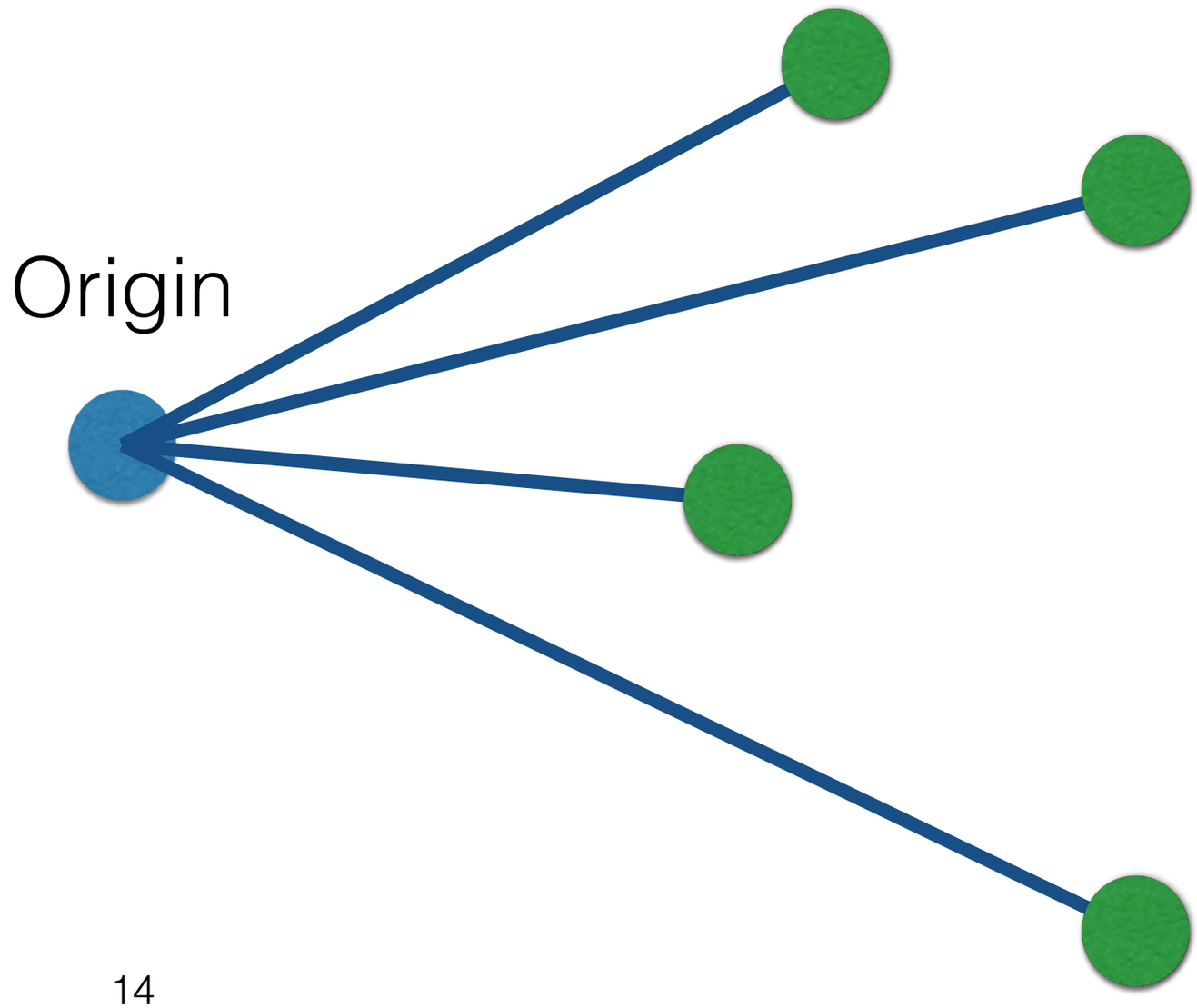
Accuracy

1) Measure ping time

2) Construct chains

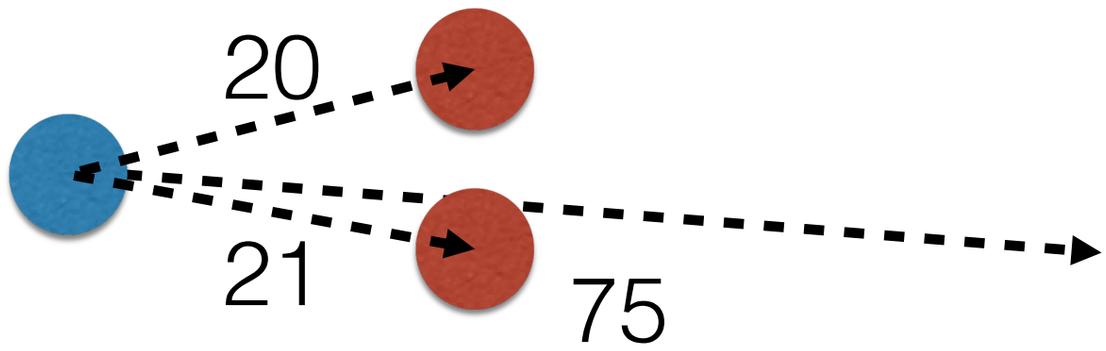
3) Compare pairs

4) Count matches



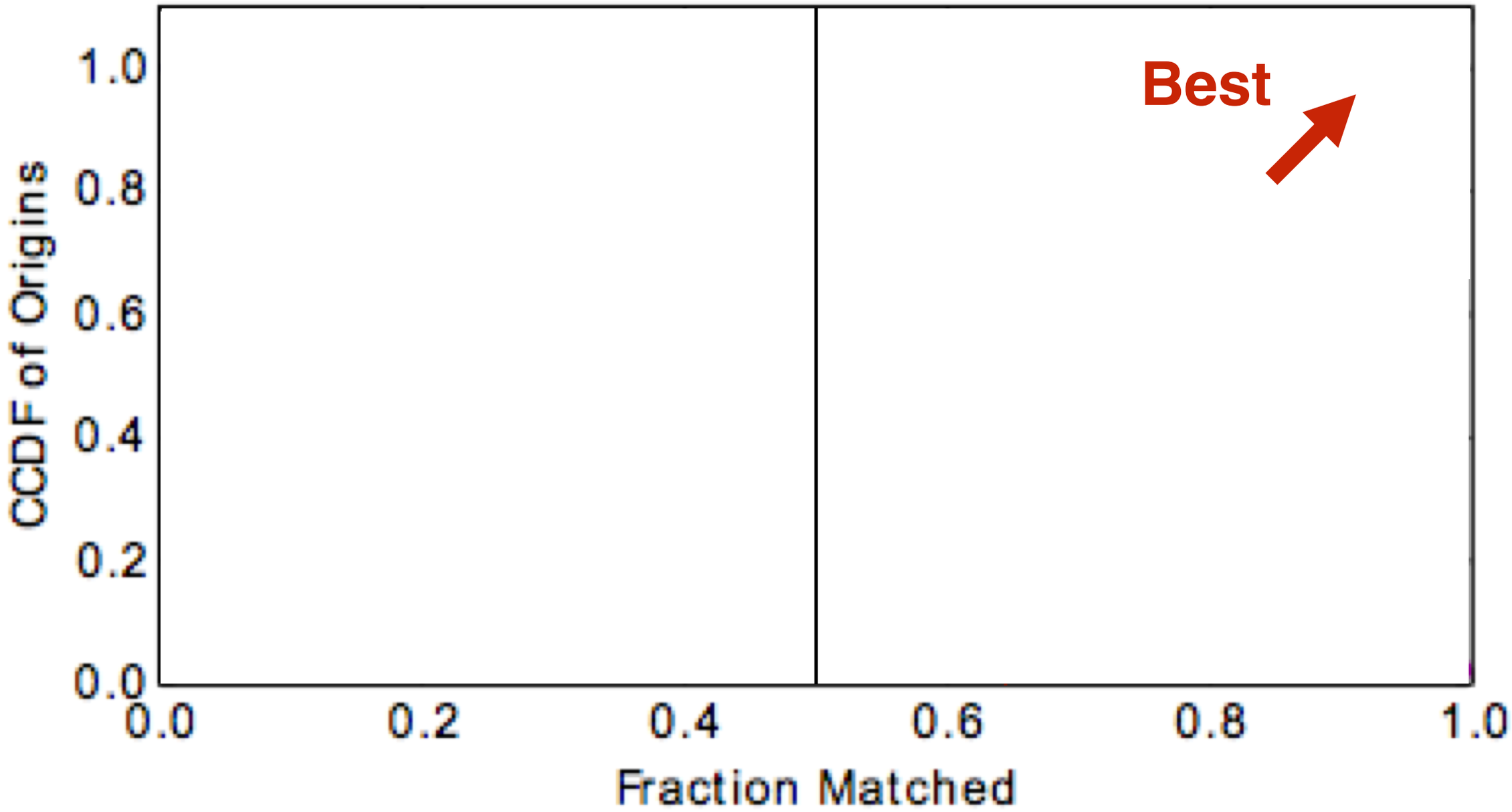
Accuracy

- What about very similar distances?



- Filter by minimum difference:
  - All, 25, 50, 100ms differences.

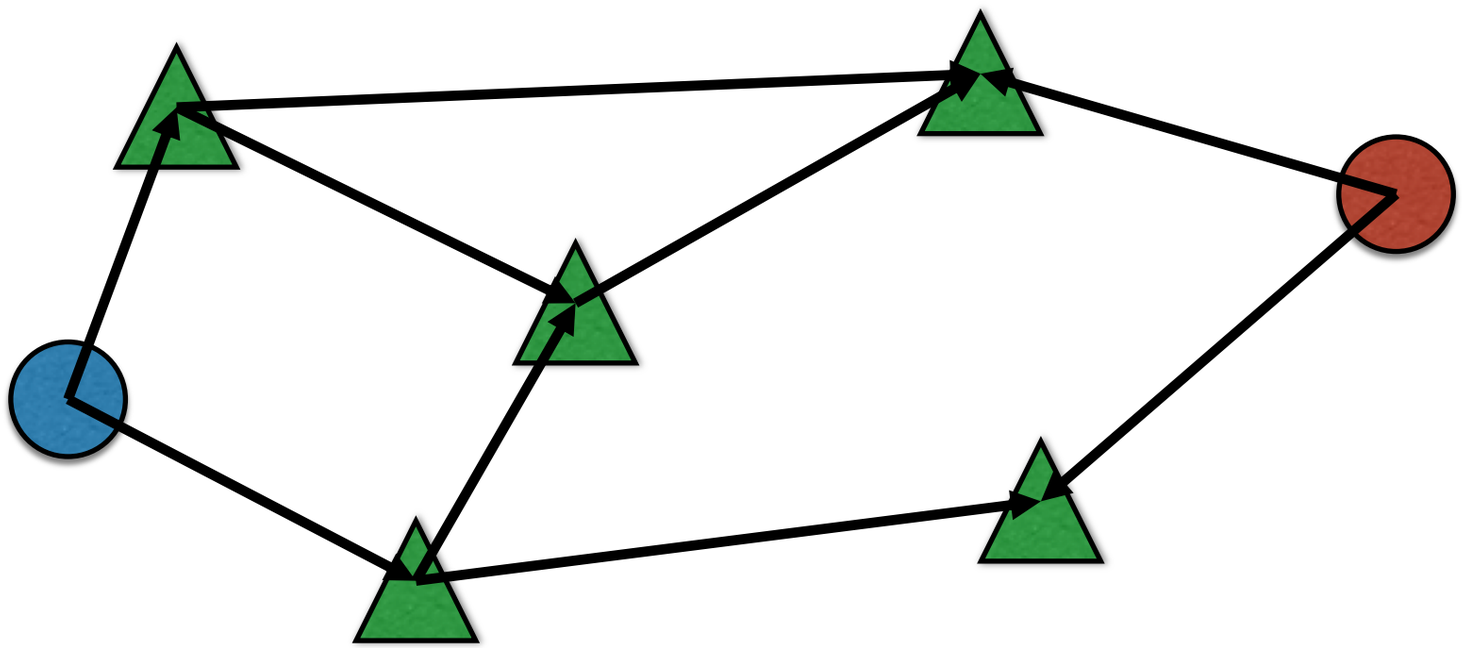
Accuracy



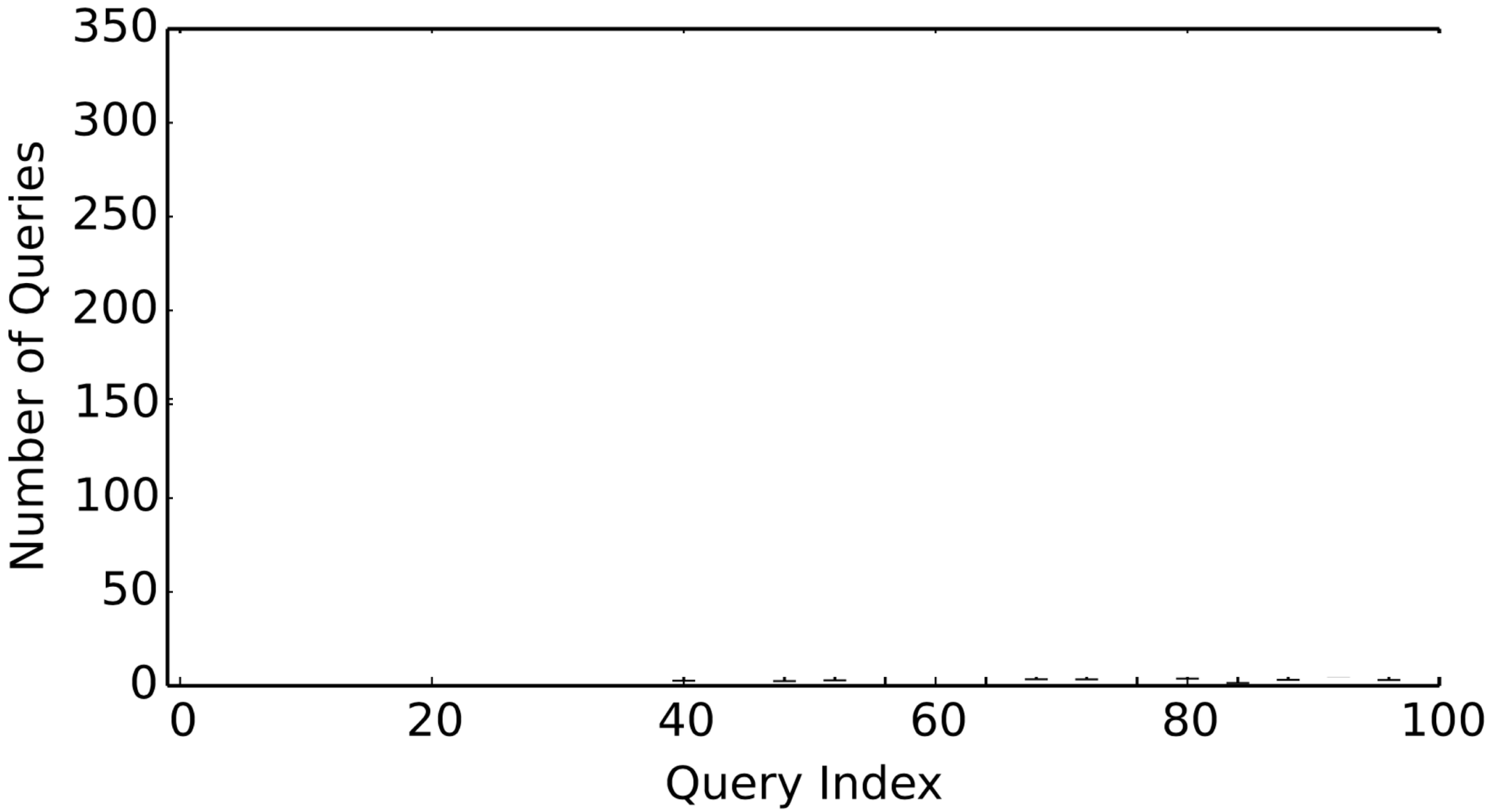
Median 83% correct

Caching

- Would like to build a consistent graph that expands as we see more hosts.
- How significantly does this graph actually reduce the queries?



Caching



After 20 paths, chains complete with under 50 queries

Summary

- Developed a system which provides an estimate of relative network distance.
- Correctly determines 83% of relative distances in the median case.
- After a cache is established, the majority of chains can be built with under 50 queries.

Thank you!