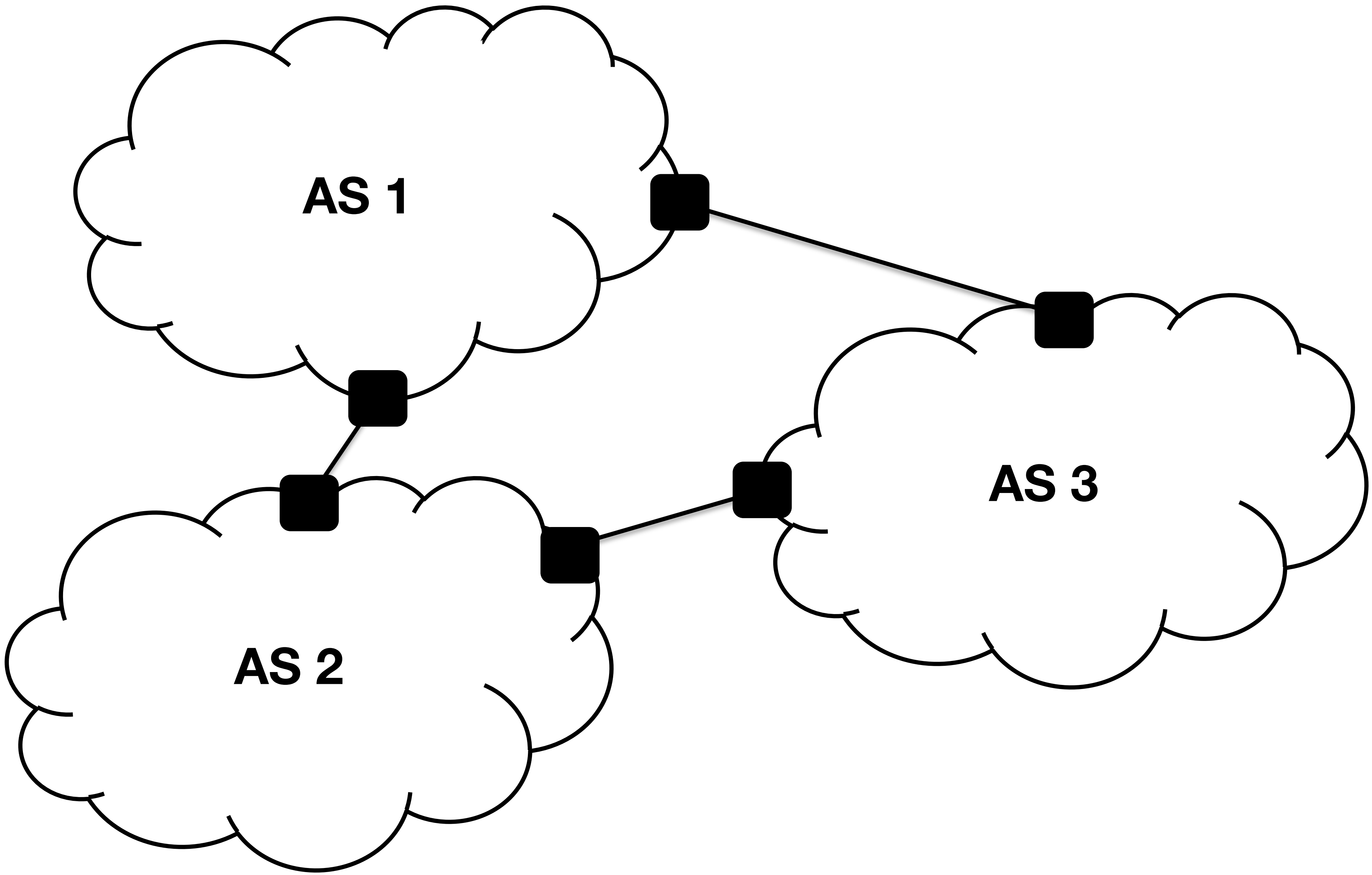
Routing

Are Internet routes stable?
symmetric?
efficient?


AS 1
AS 3
AS 2

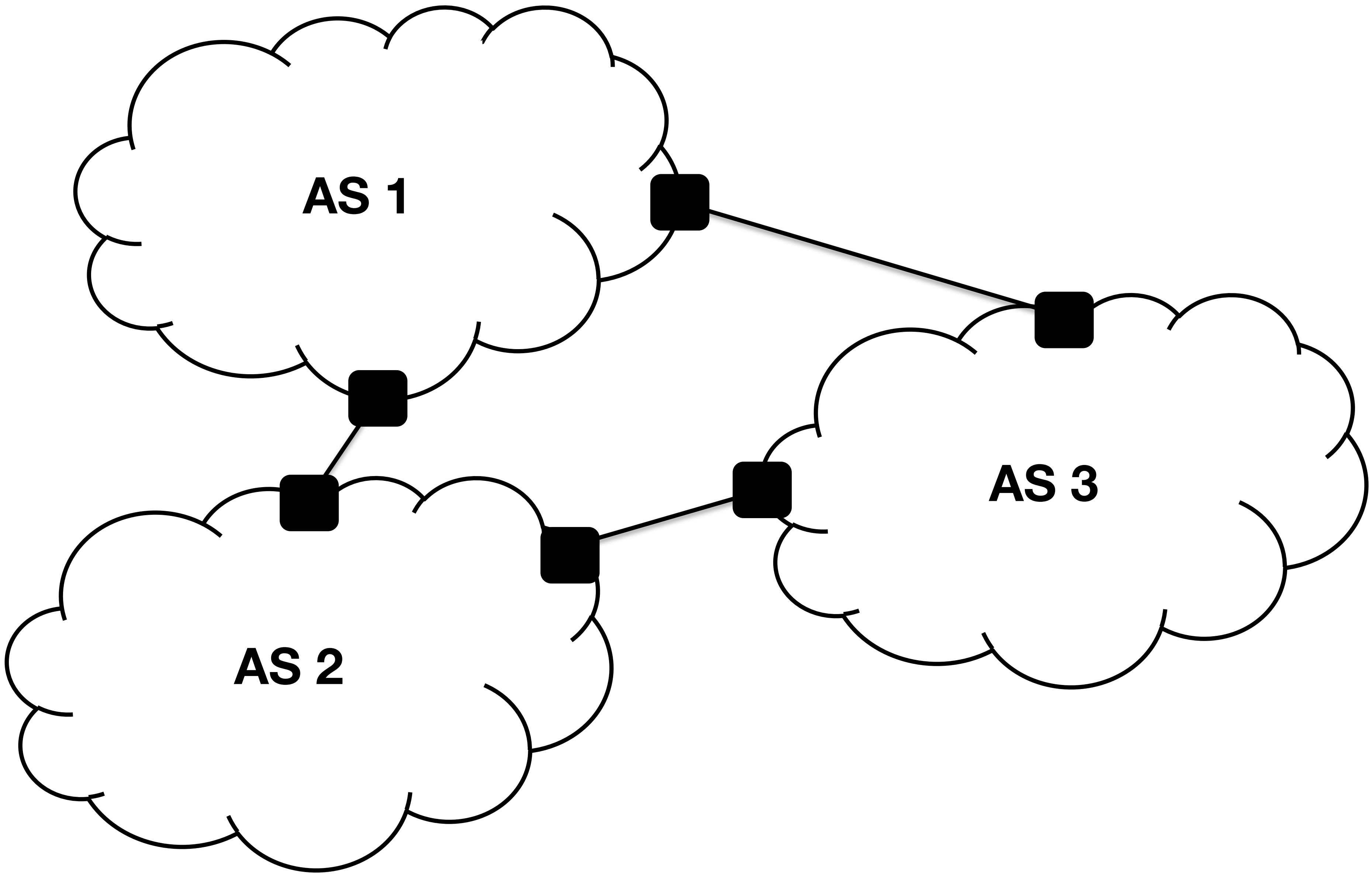
Intra-Domain Routing

ISPs are free to use own metrics to route (e.g. hop count)

Inter-Domain Routing (using BGP)

Depends on policy:
business contract,
load balancing
quality of routes

“early exit” routing (or hot potato routing)


AS 1
AS 2
AS 3

“End-to-End Routing Behavior in the Internet”

V. Paxson

SIGCOMM 96

(2006 SIGCOMM Test of Time Award)

37

hosts

mean inter-measurement interval of 2 hours & 2.75 days

traceroute

measure A to B immediately after B to A

~50%

of ASs

(weighted by their importance)

Routing Pathologies

Routing Loops

```
1 ir6gw.lbl.gov 1.853 ms 1.623 ms 2.358 ms
2 er1gw.lbl.gov 7.165 ms 2.996 ms 3.098 ms
3 ir2gw.lbl.gov 4.882 ms 3.516 ms 8.371 ms
4 isdn1gw.lbl.gov 7.98 ms 4.393 ms 4.311 ms
5 ascend49.lbl.gov 36.833 ms 32.772 ms 31.428 ms
6 isdn1gw.lbl.gov 30.428 ms 30.502 ms 33.528 ms
7 ascend49.lbl.gov 69.006 ms 59.429 ms 58.82 ms
8 isdn1gw.lbl.gov 59.358 ms 63.734 ms 61.775 ms
9 ascend49.lbl.gov 85.629 ms 84.168 ms 83.397 ms
10 isdn1gw.lbl.gov 83.374 ms 83.201 ms 83.349 ms
11 ascend49.lbl.gov 110.316 ms 120.243 ms 116.84 ms
12 isdn1gw.lbl.gov 109.221 ms 108.97 ms 109.242 ms
13 ascend49.lbl.gov 135.867 ms 136.797 ms 140.849 ms
     :
     :
```

50

occurrences of loops

loops can last for hours

clustered geographically and temporally

confined within one AS

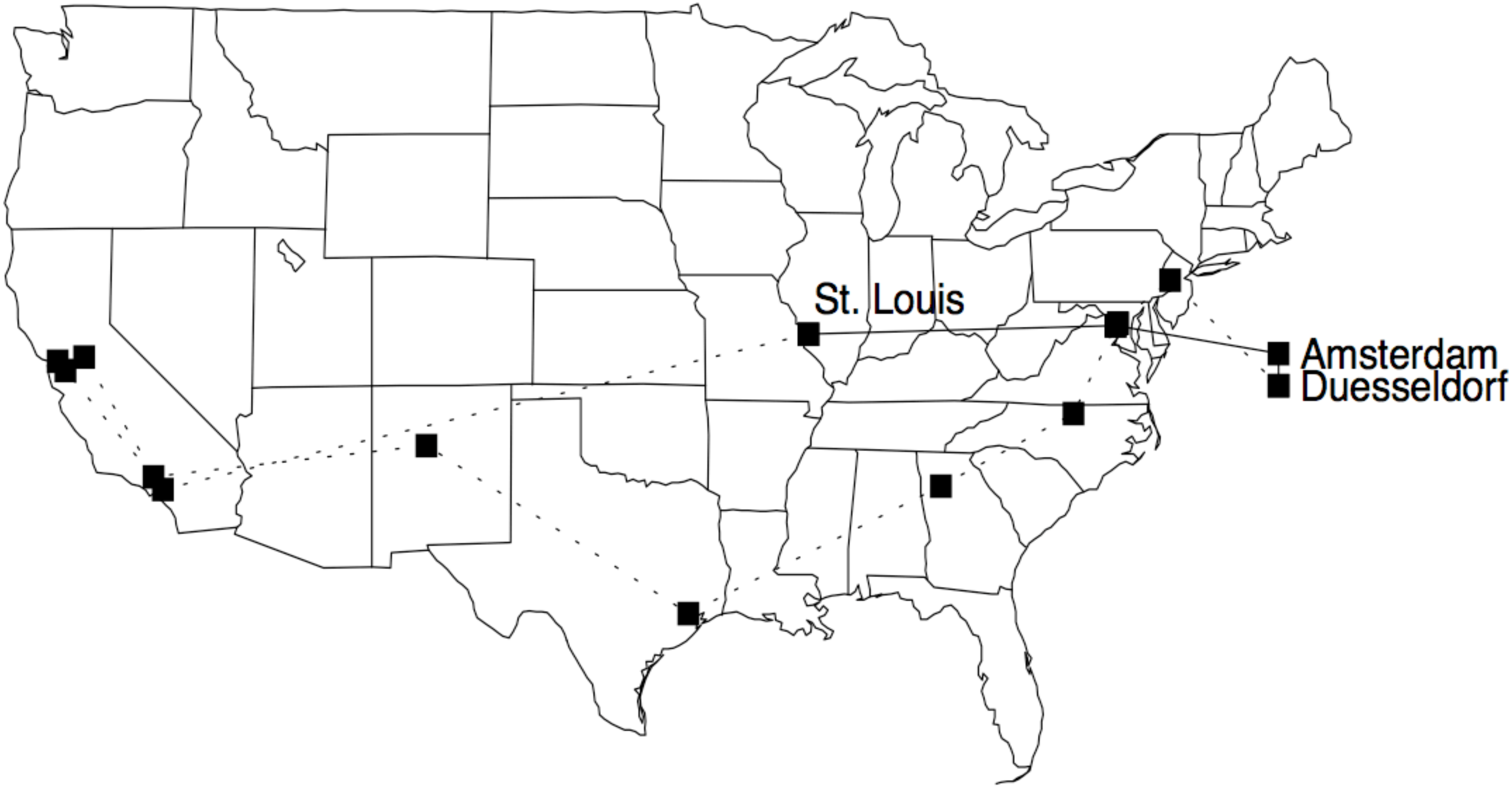
Routing Error

1 mfd-01.rt.connix.net 8 ms 4 ms 3 ms
2 sl-dc-5-s2/0-512k.sprintlink.net 39 ms 39 ms 39 ms 3 sl-dc-6-f0/0.sprintlink.net 39 ms 38 ms 50 ms
4 psi-mae-east-1.psi.net 48 ms 66 ms *
5 * * core.net218.psi.net 90 ms
6 192.91.187.2 1139 ms 1188 ms *
7 * * *
8 biu-tau.ac.il 1389 ms * *
9 tau.man.ac.il 1019 ms * *
10 * * *
11 * cisco301s1.huji.ac.il 1976 ms *
12 * * *
13 * * *
14 * * cisco101e5.huji.ac.il 1974 ms
15 * * *
16 * cisco103e2.gr.huji.ac.il 1010 ms 1069 ms
17 cisco101e01.cc.huji.ac.il 2132 ms * *
18 cisco102e13.huji.ac.il 888 ms 976 ms 2005 ms
19 cisco103e2.gr.huji.ac.il 1657 ms * *

A route to London ends up in Israel?

Route Fluttering

```
1 fpls.postech.ac.kr 2 ms 2 ms 2 ms
2 fddicc.postech.ac.kr 3 ms 2 ms 2 ms
3 ktrc-postech.hana.nm.kr 57 ms 123 ms 30 ms
4 gateway.hana.nm.kr 31 ms 31 ms 31 ms
5 hana.hana.nm.kr 33 ms 140 ms 32 ms
6 bloodyrouter.hawaii.net 825 ms 722 ms 805 ms
7 usa-serial.gw.au 960 ms 922 ms 893 ms
8 national-aix-us.gw.au 1039 ms * *
9 * rb1.rtr.unimelb.edu.au 903 ms rb2.rtr.unimelb.edu.au 1279 ms
10 itee.rtr.unimelb.edu.au 1067 ms 1097 ms 872 ms
11 * * mulkirri.cs.mu.oz.au 1468 ms
12 mullala.cs.mu.oz.au 1042 ms 1140 ms 1262 ms
```



Taken from Paxson's PhD Thesis: Alternate routes are taken for packets from WUSTL to U Mannheim

Are Routes Stable?

are network paths predictable?

end-to-end measurement: the same path?

prevalence

"given a route r observed at present, how likely to observe r again in future?"

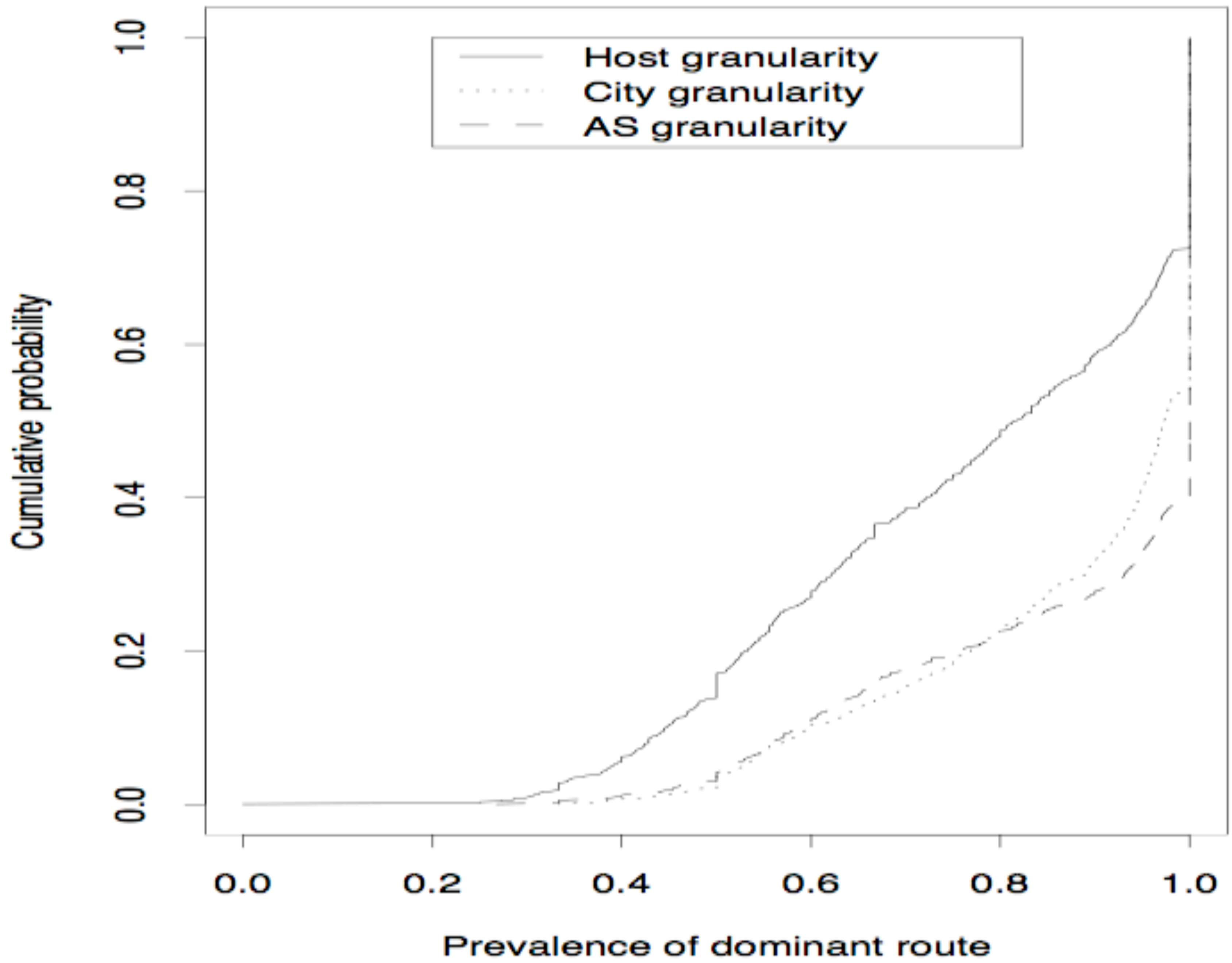
persistence

"given a route r observed at time t, how long before this route is likely to have changed?"

1, 1, 1, 1, 2, 1, 1, 1, 1

prevalence of r = k/n

we make n traceroute and k of them shows route r



Host granularity
City granularity
AS granularity
Cumulative probability
0.0
0.2
0.4
0.6
0.8
1.0
Prevalence of dominant route

in one instance, two sites exhibit 9 diff routes

At UCL, the prevalence of dominant routes is below 0.5

In general, paths are dominated by a single route, but there is significant site-to-site variations.

persistence

"given a route r observed at time t, how long before this route is likely to have changed?"

Summary: occur over large time scale

seconds and minutes: flutter and tightly coupled routers

10s of minutes:
9%

hours:
intra network changes
(4%)

6+ hours:
intra network changes
(19%)

68% shows persistence over days

Are Routes Symmetric?

49%

of routes are asymmetric

(>1 diff city)

30%

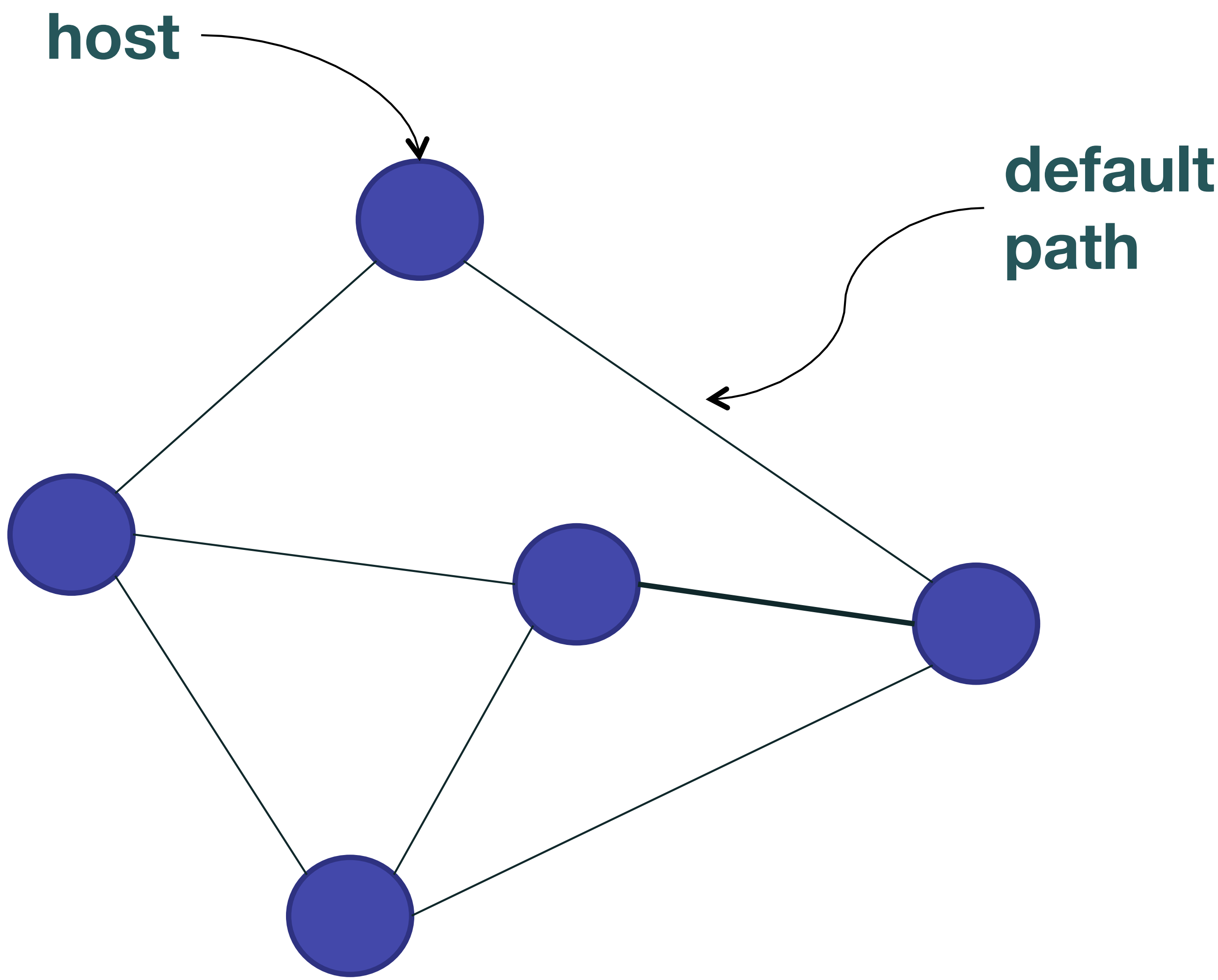
of routes are asymmetric

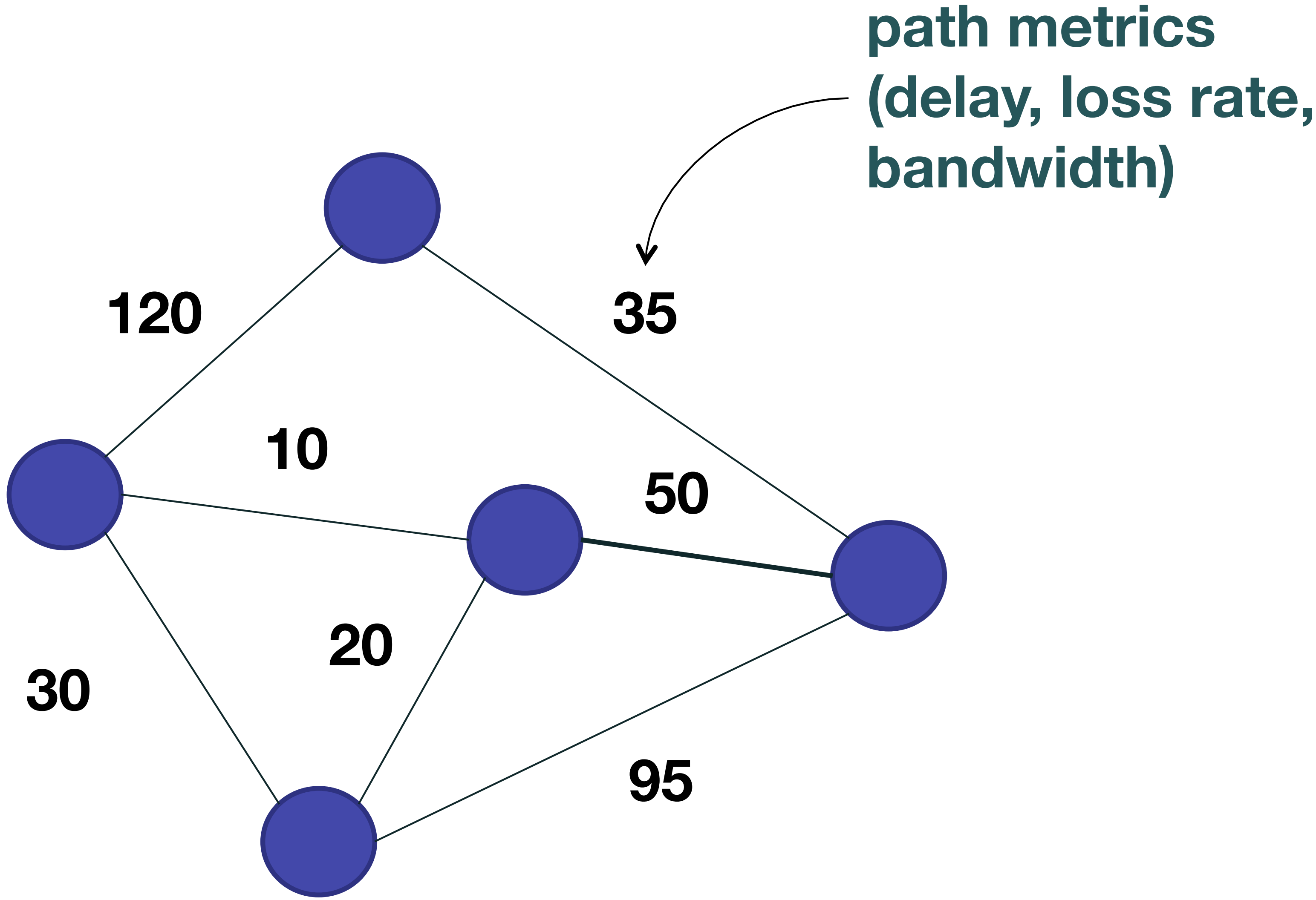
(>1 diff AS)

Are Routes Optimal?

“The End-to-End Effects of Internet Path Selection”

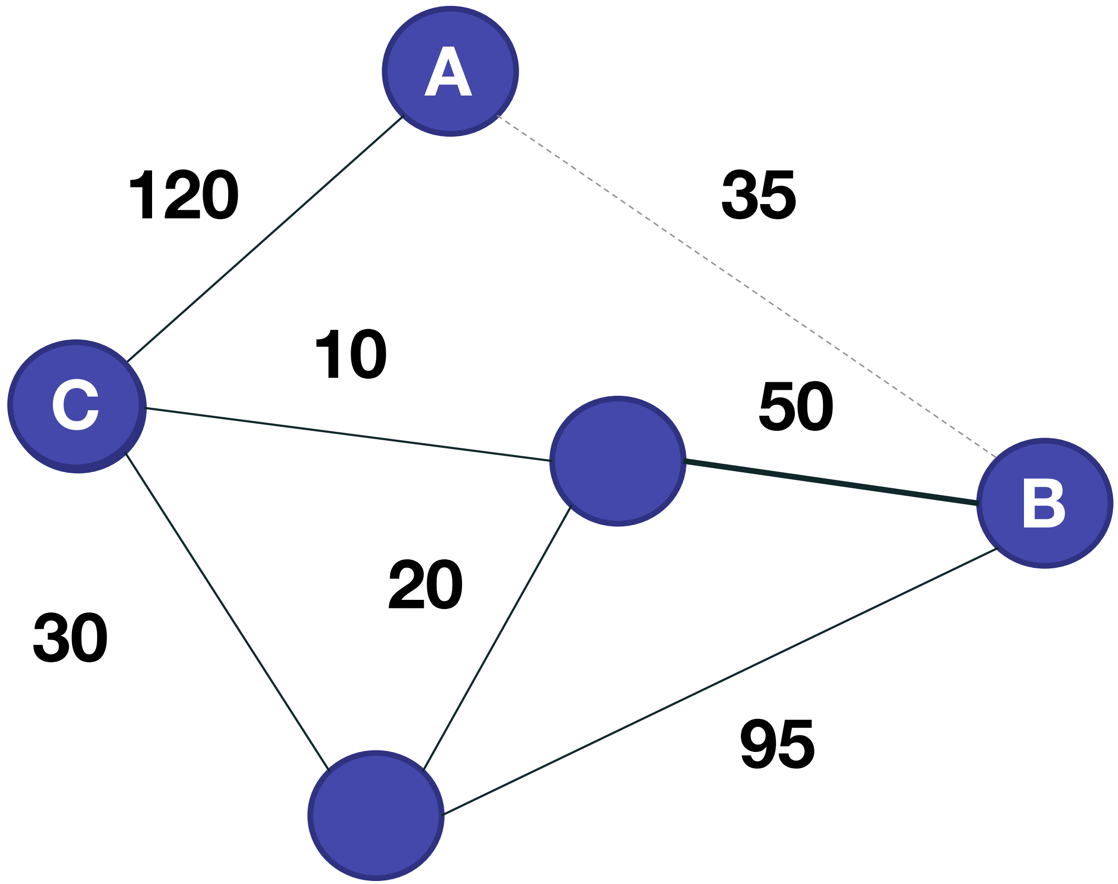
S. Savage et al.


host
default
path


path metrics
(delay, loss rate,
bandwidth)
120
35
10
50
20
30
95

Uses Paxson’s Dataset
+ 3 New Sets
(UW1, UW3, UW4)

Can we find alt path with shorter RTT?


A
120
35
10
C
50
B
20
30
95

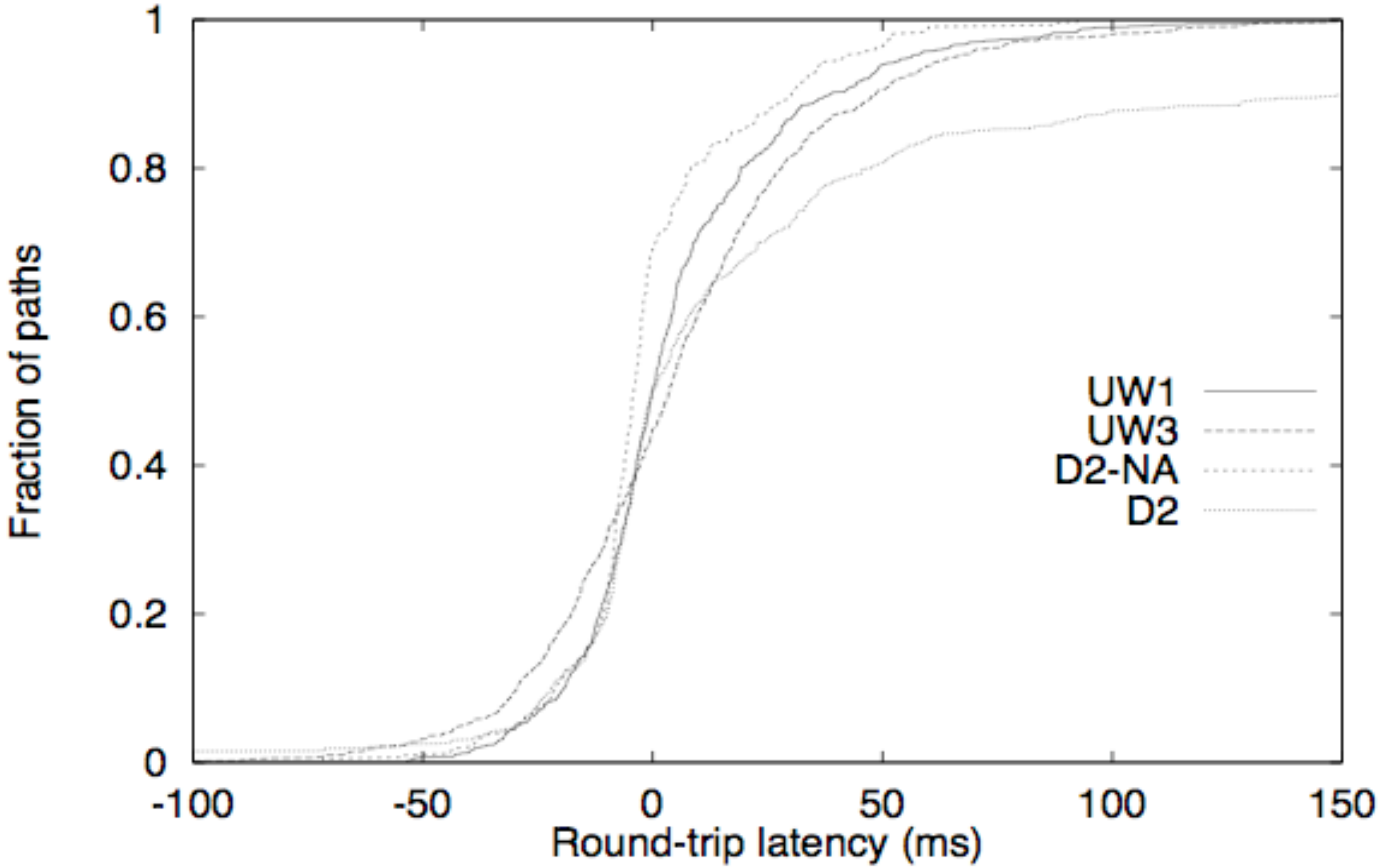
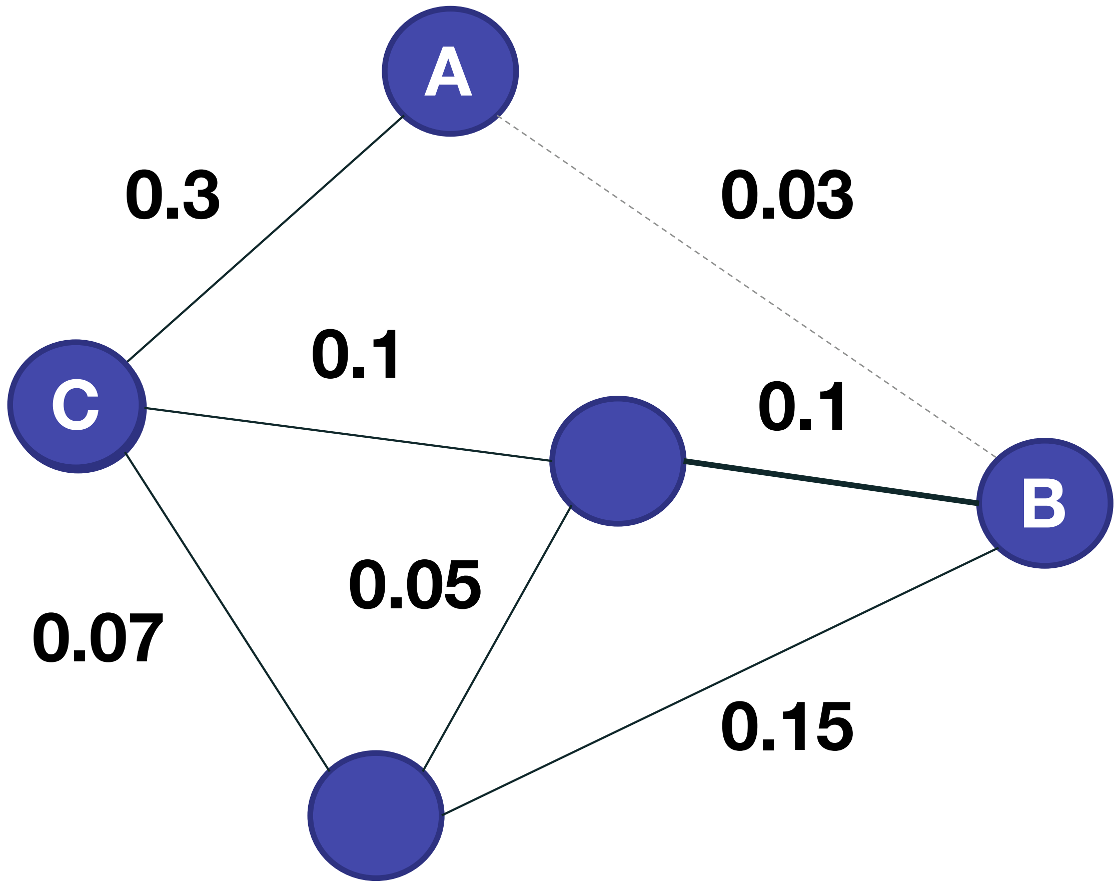


Figure 1: CDF of the difference between the mean round-trip time recorded on each path, and the best mean round-trip time derived for an alternate path.

Can we find alt path with lower loss rate?

Assuming that losses are uncorrelated



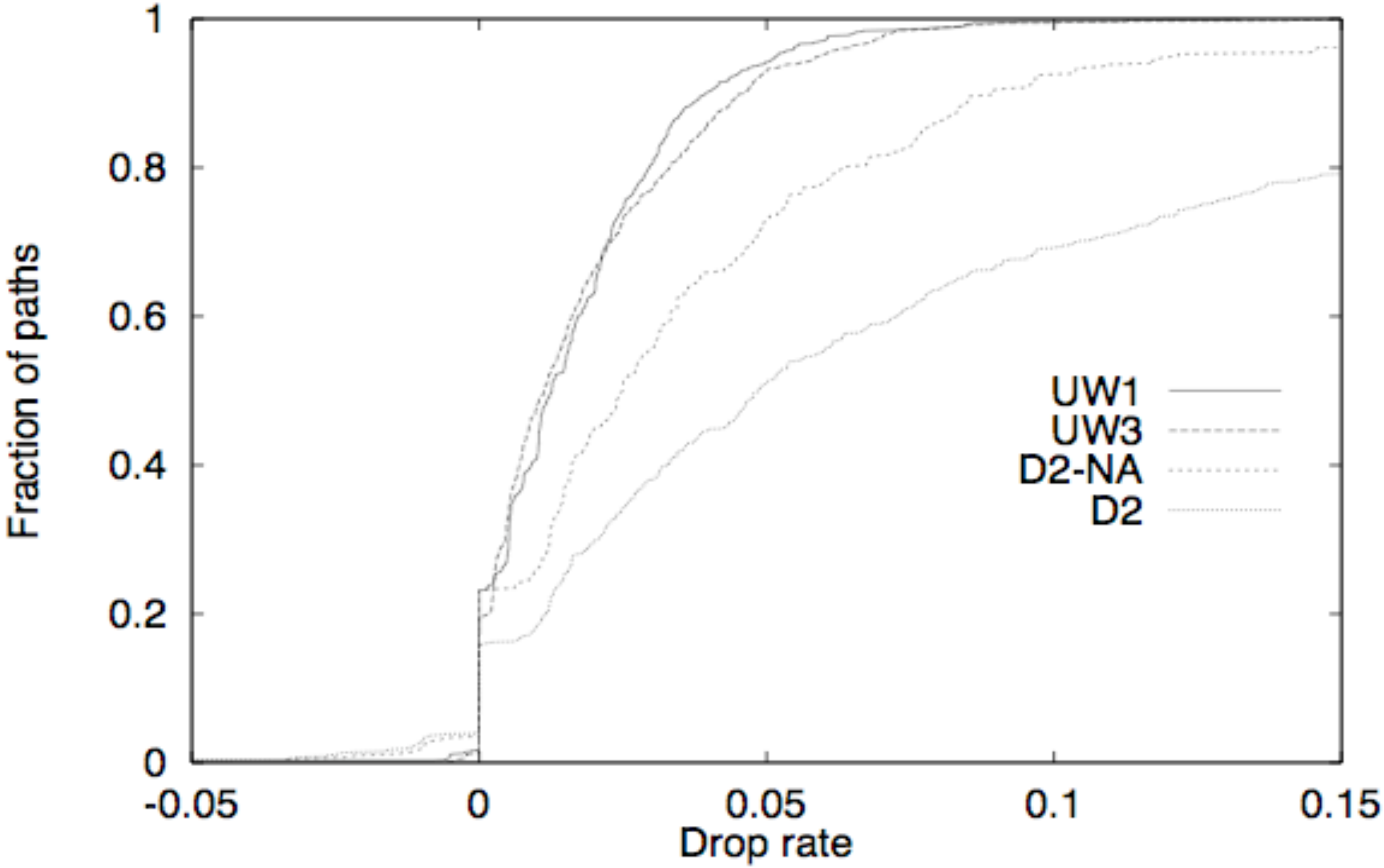
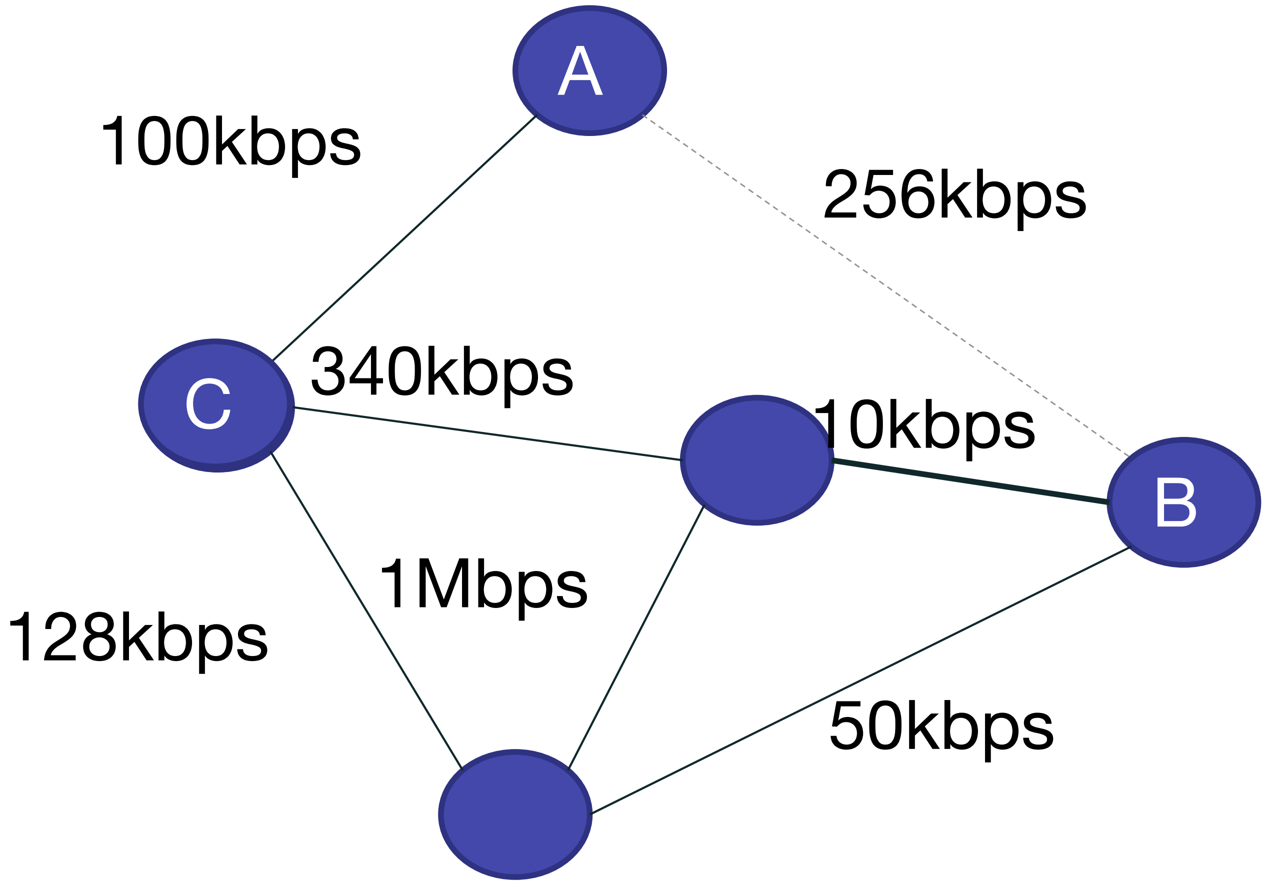
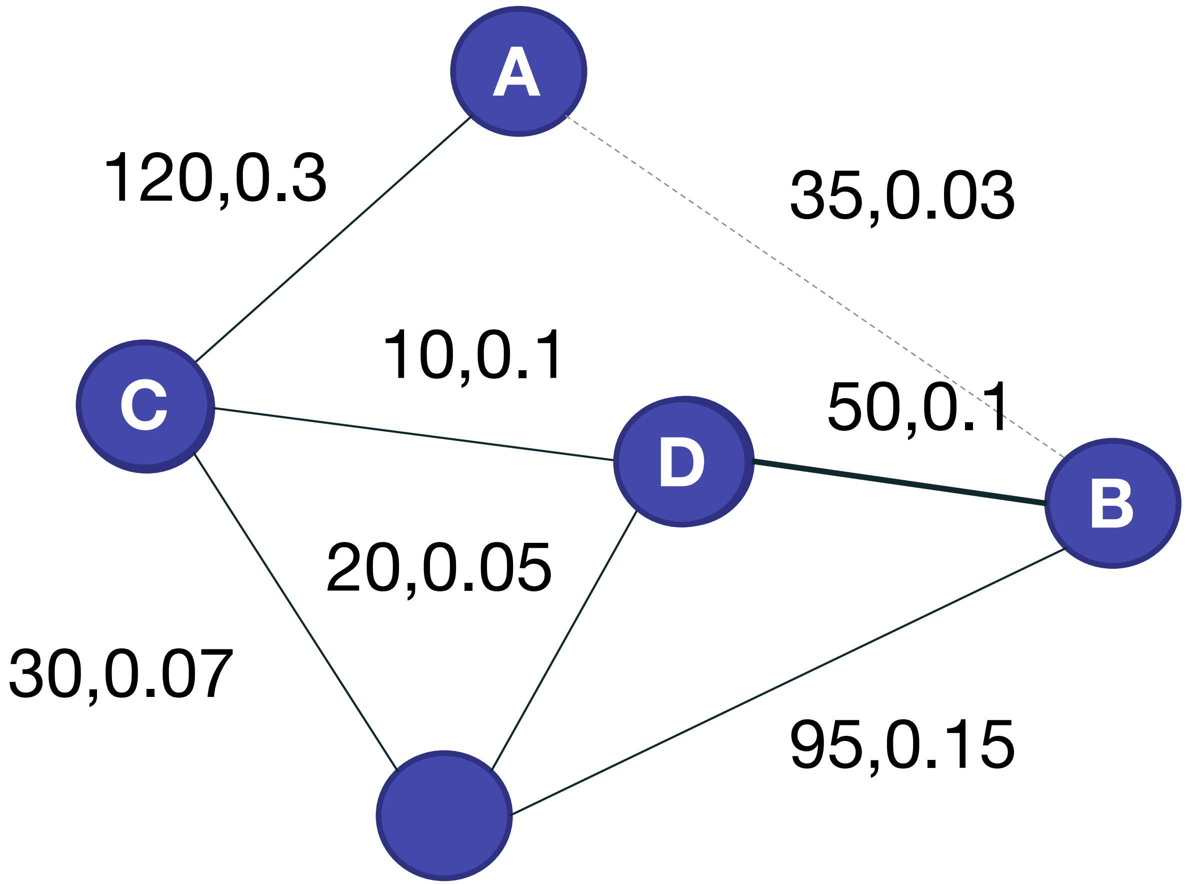


Figure 3: CDF of the difference between the mean loss rate recorded on each path, and the best mean loss rate derived for an alternate path.

Can we find alt path with higher bandwidth?

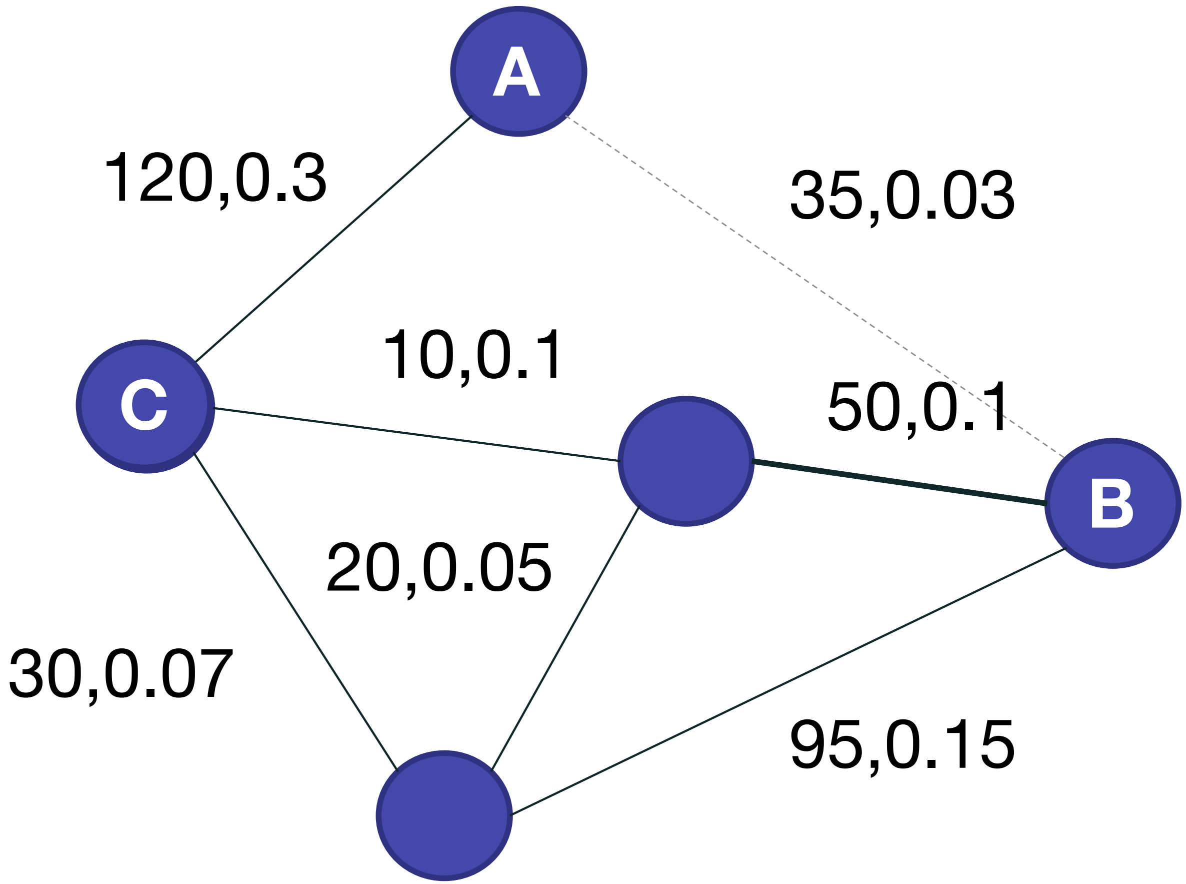

A
100kbps
256kbps
C
340kbps
10kbps
B
1Mbps
128kbps
50kbps

optimistic model



loss rate ACDB = 0.3

pessimistic model



loss rate ACDB = 0.5

$$\frac{MSS}{RTT}\sqrt{\frac{3}{2p}}$$

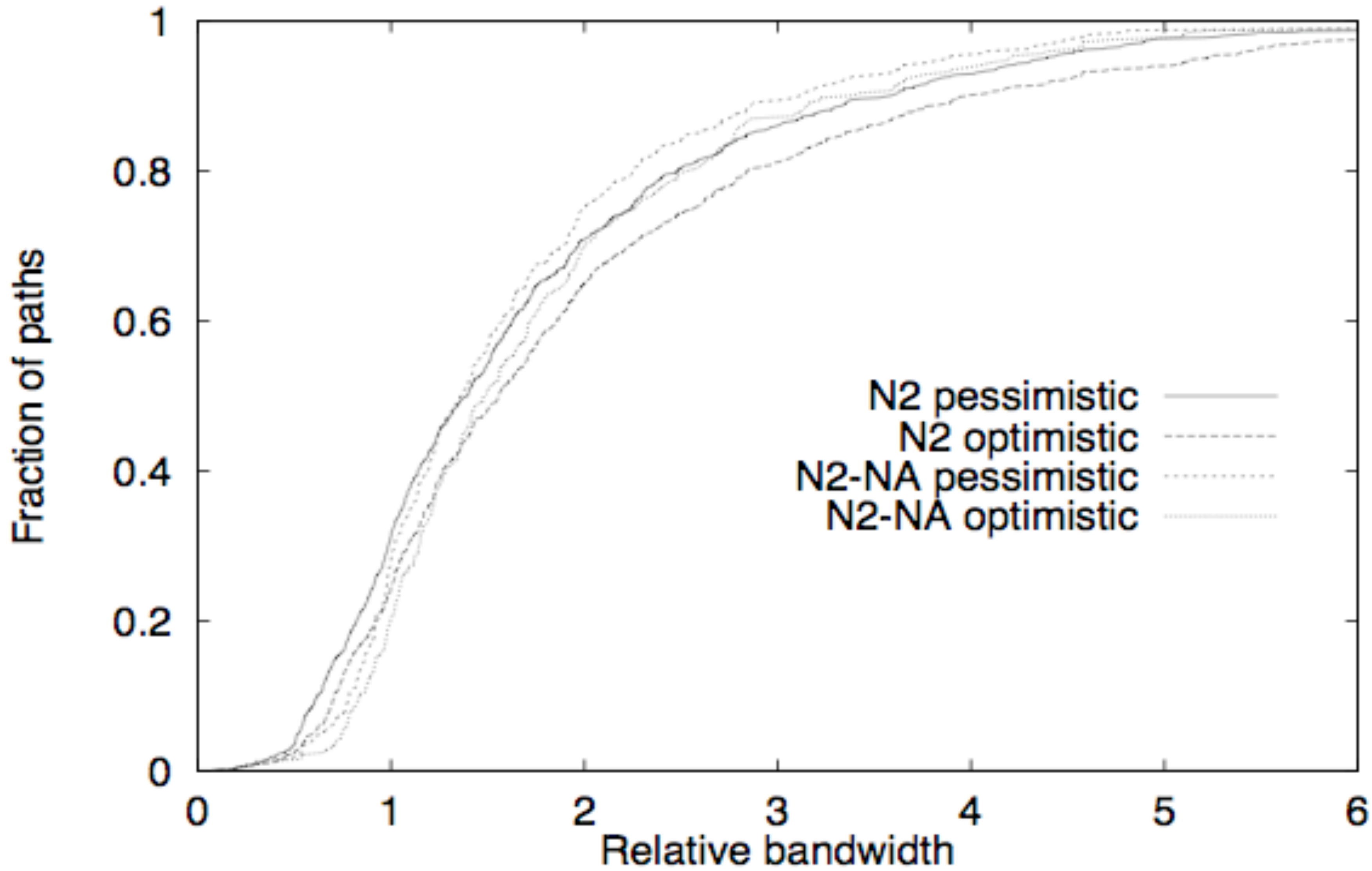
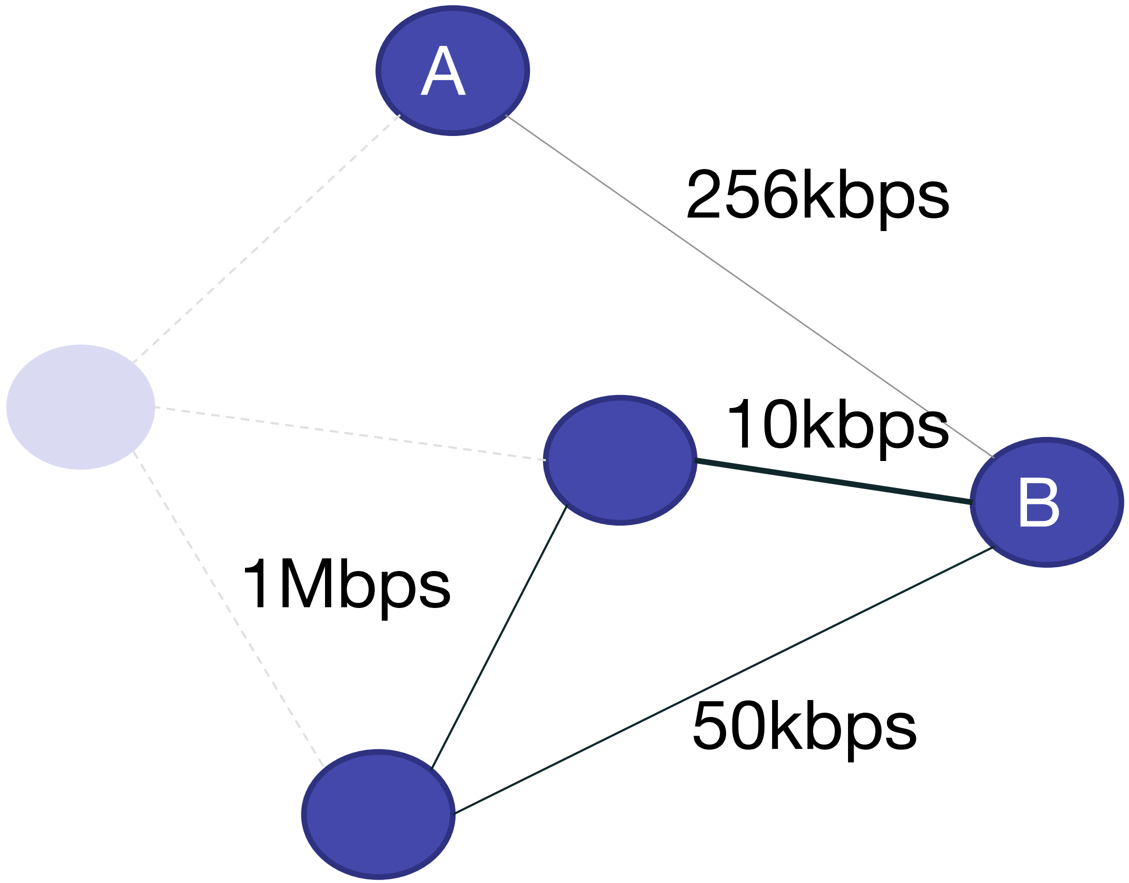


Figure 5: CDF of the *ratio* between the mean bandwidth recorded on each path, and the best mean bandwidth derived for a one-hop alternate path. The lines labeled “optimistic” and “pessimistic” reflect the same two cases as in Figure 4

Are these better alt paths due to a small set of hosts?

For each node, remove from graph, repeat experiments.
Find 10 nodes which affected the results the most.



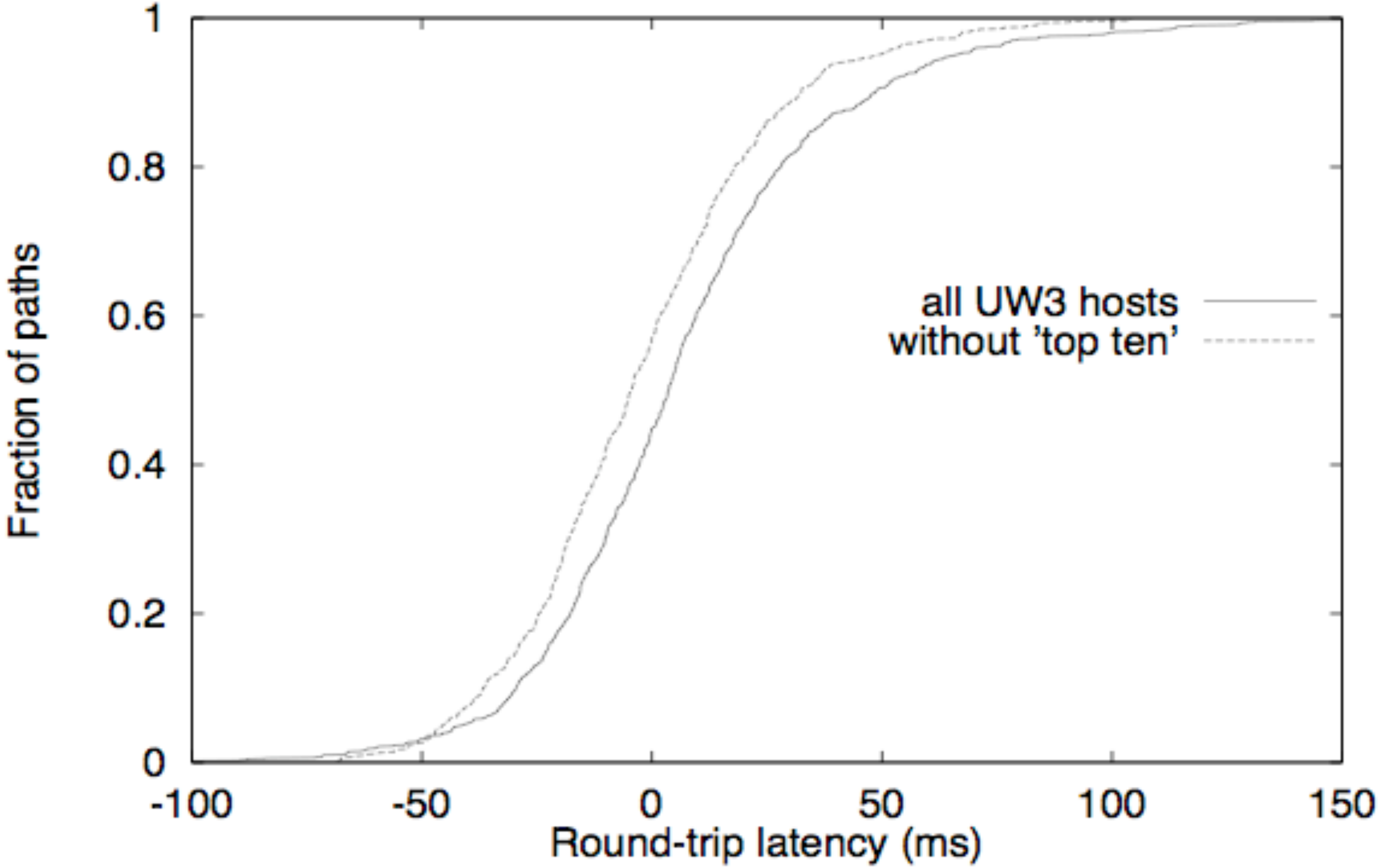


Figure 12: CDF of the difference between the mean round-trip time recorded on each path and the best mean round-trip time derived for an alternate path, and the equivalent CDF computed for the dataset after having removed the "top ten" hosts. This graph is for the UW3 dataset.

For each node, find how many times it appears as an intermediate host in some superior alternate path.

A	3	C-A-B, D-A-E-B, F-A-G-C
B	2	A-B-C, D-B-A-E
:	:	:

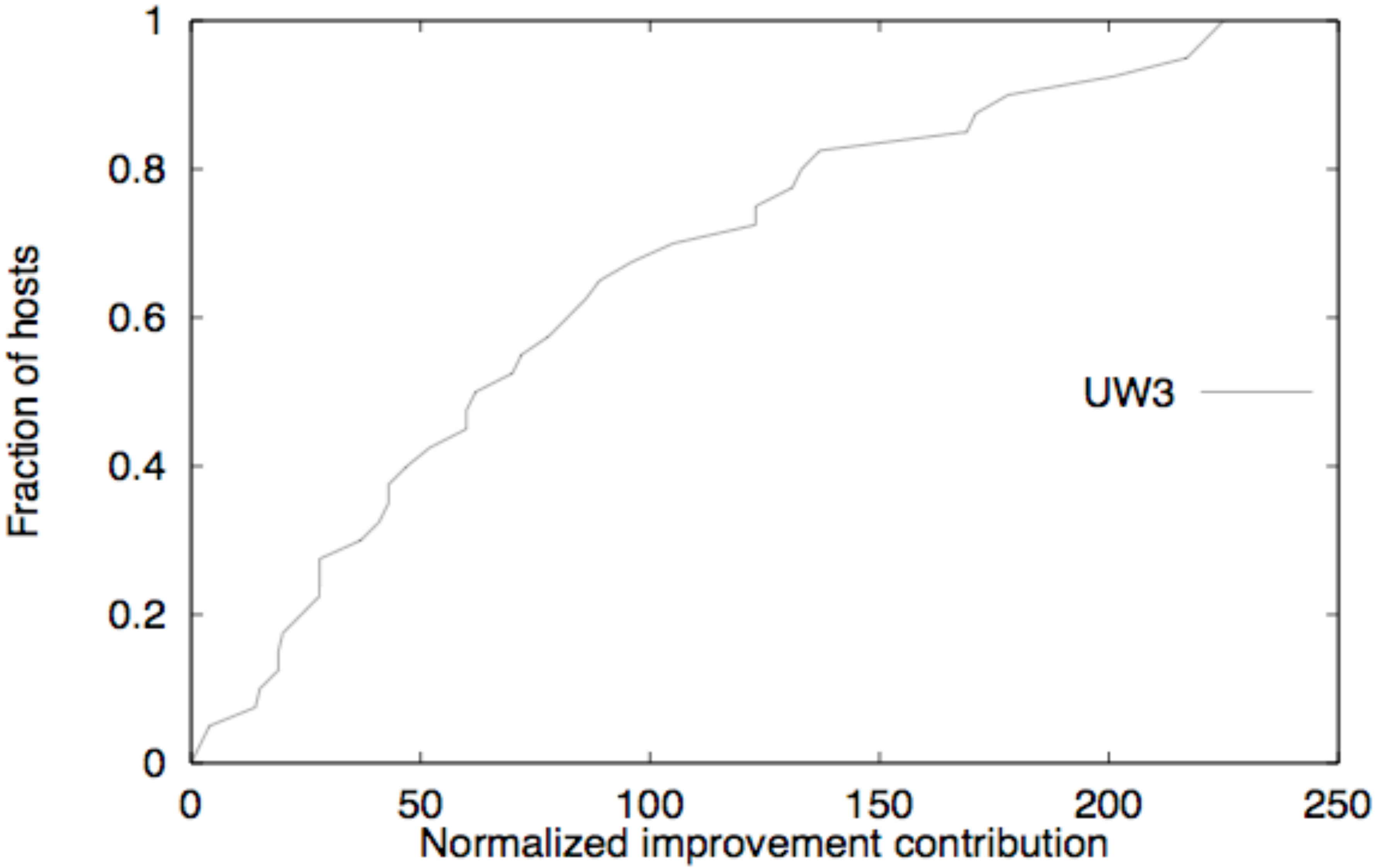
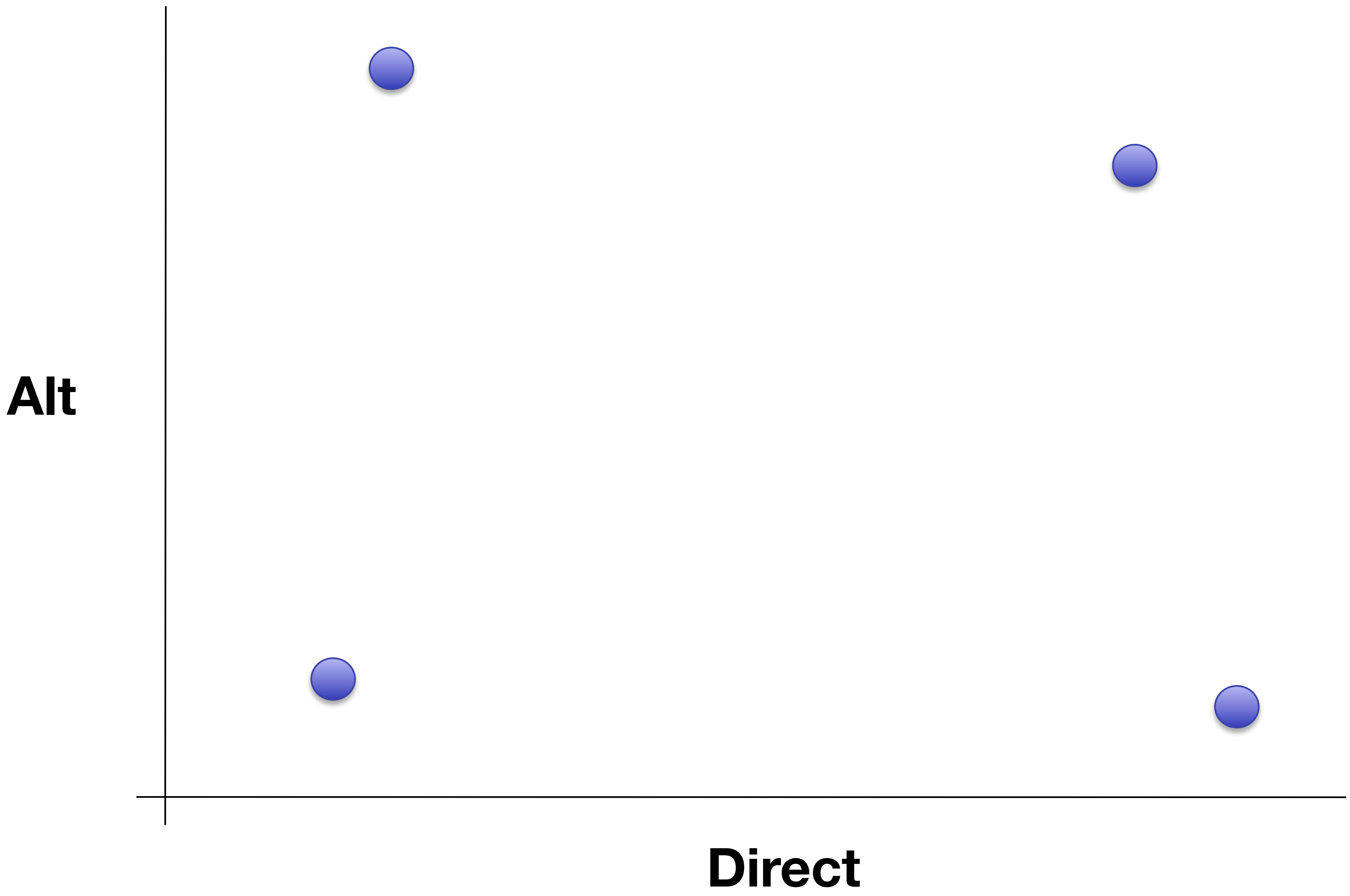


Figure 13: CDF of the number of better alternate paths in which a host appears as an intermediate node, weighted by the degree to which the alternate path is better

Are these better alt path due to a small set of ASes?


Alt
Direct

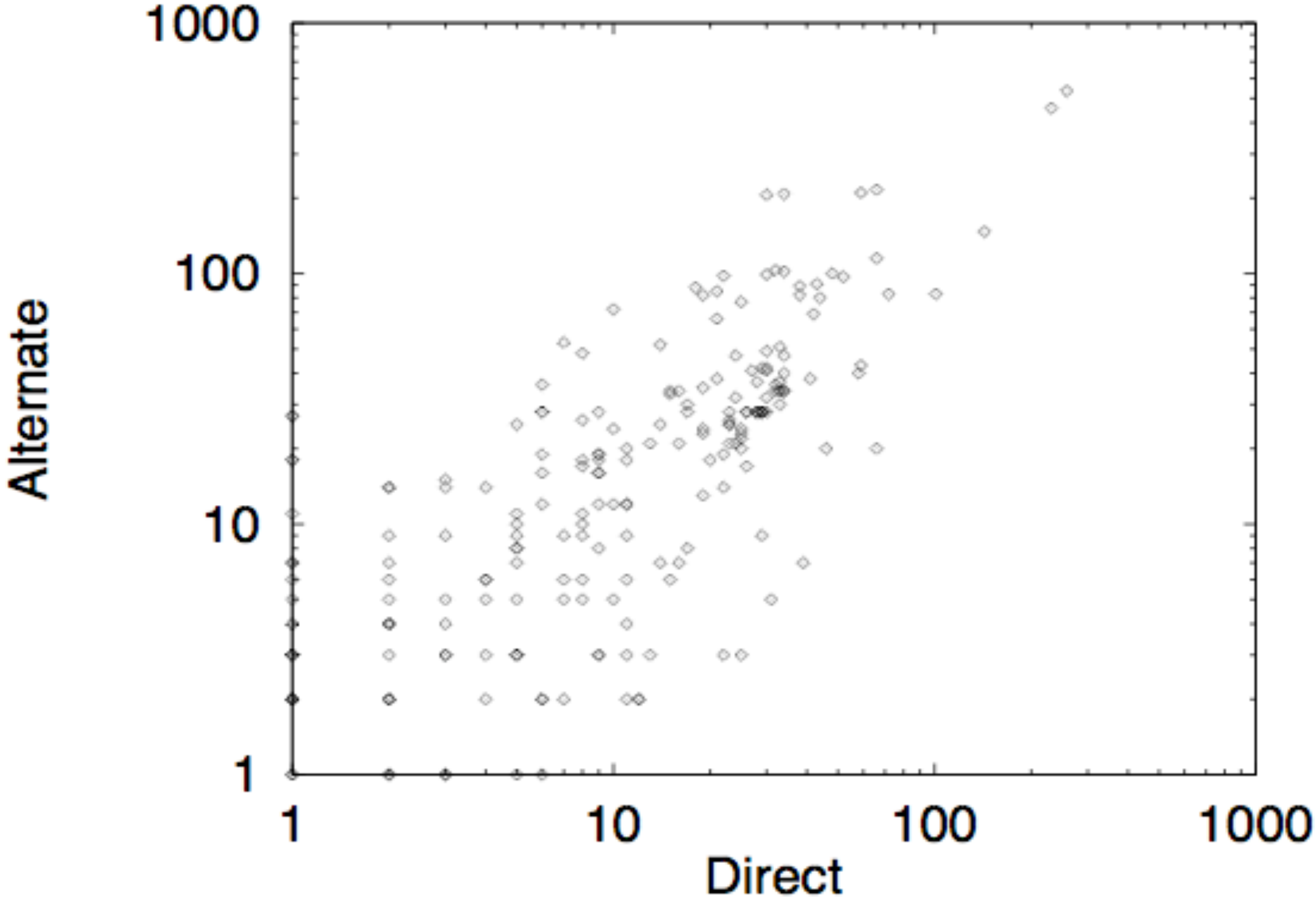


Figure 14: Scatterplot of AS's found in the UW1 dataset. The x-axis represents the number of default paths in which that AS appears, while the y-axis is the number of best alternate paths (for the metric of round-trip time) in which it appears.

Are shorter alt path due to less congestion?

Estimate prop delay as the 10%-tile delay on a path

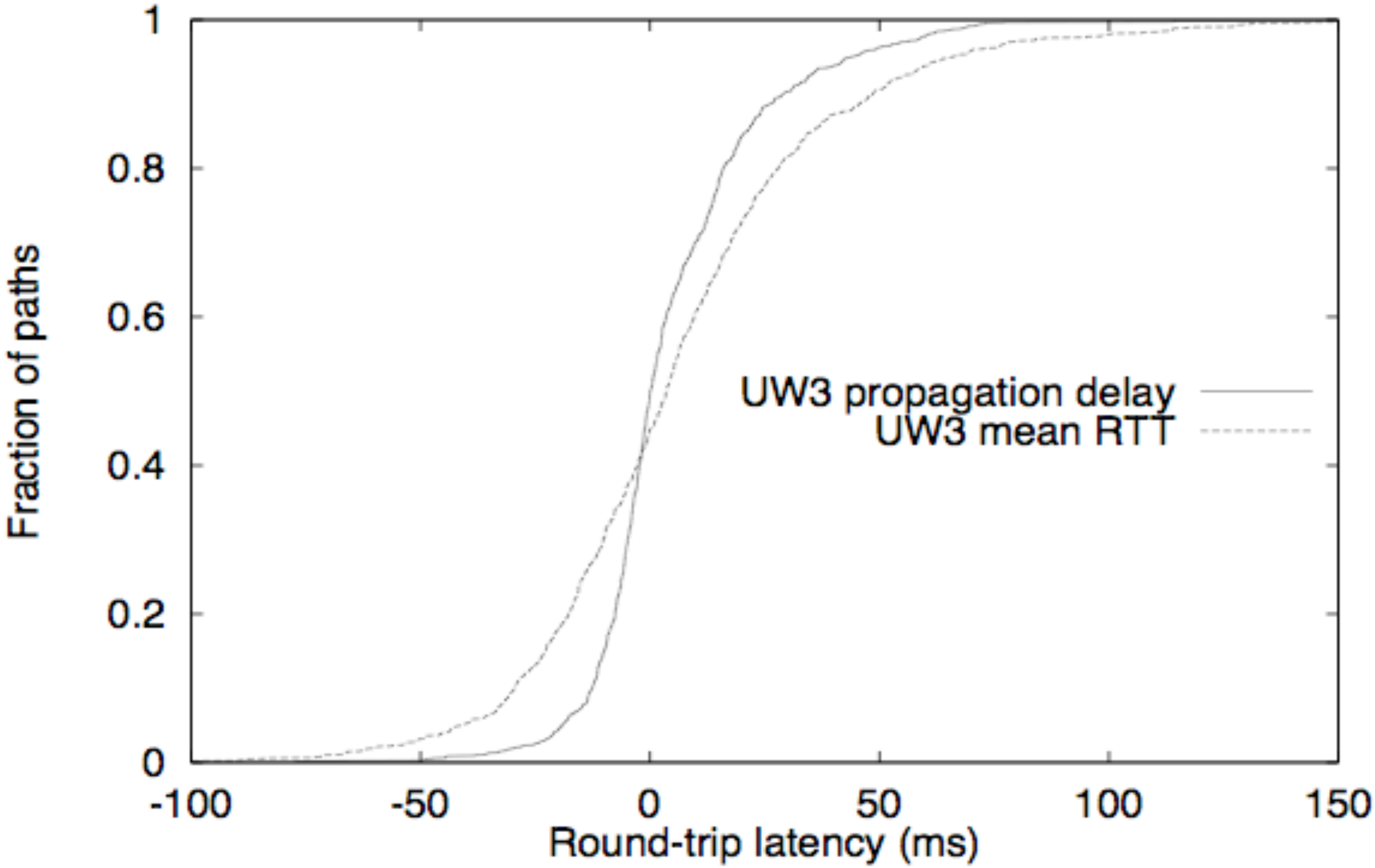
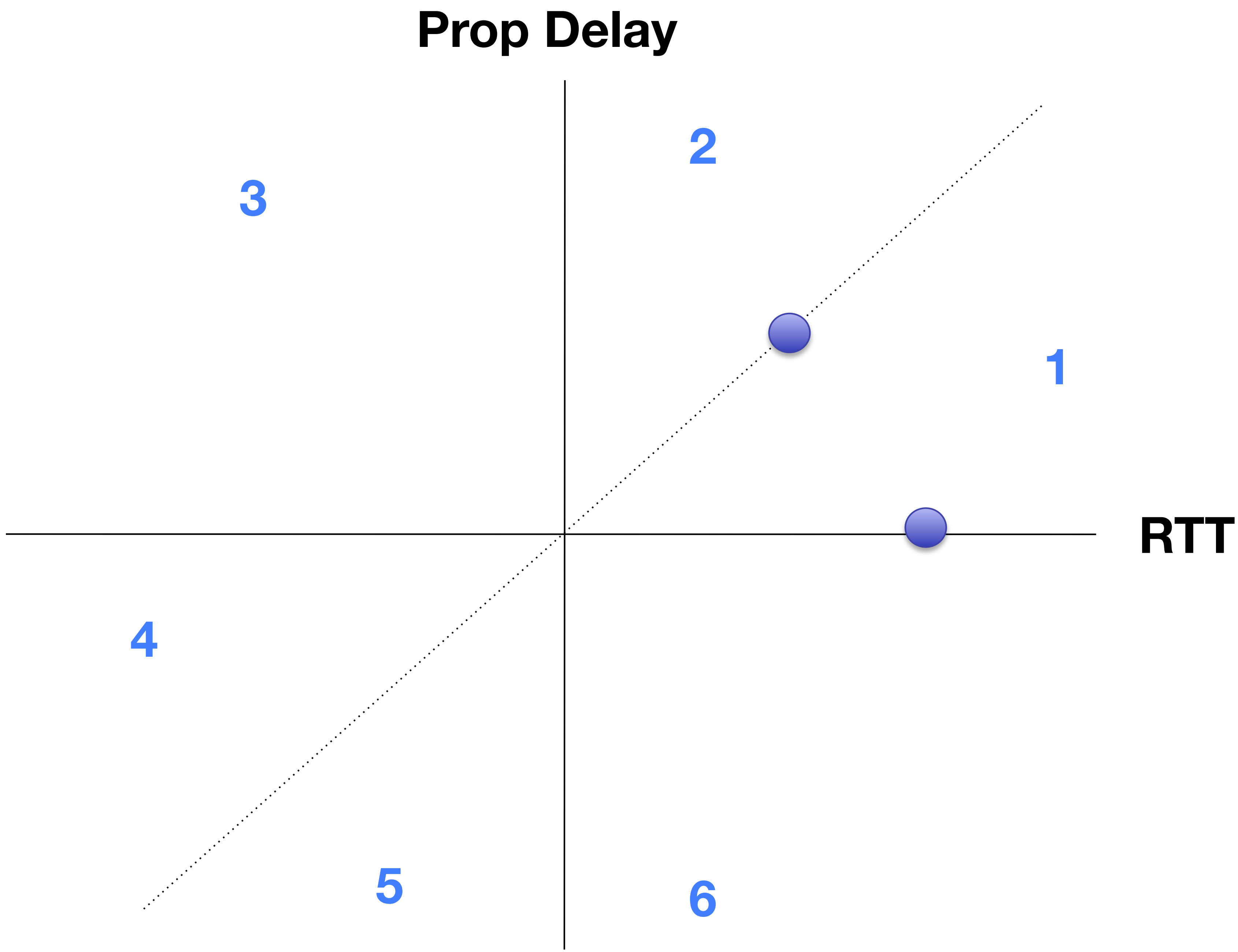


Figure 15: CDF of the difference between the propagation delay recorded for each path, and the best propagation delay derived for an alternate path, superimposed with the equivalent CDF for the mean round-trip time. This graph is for the UW3 dataset.


Prop Delay
RTT
1
2
3
4
5
6

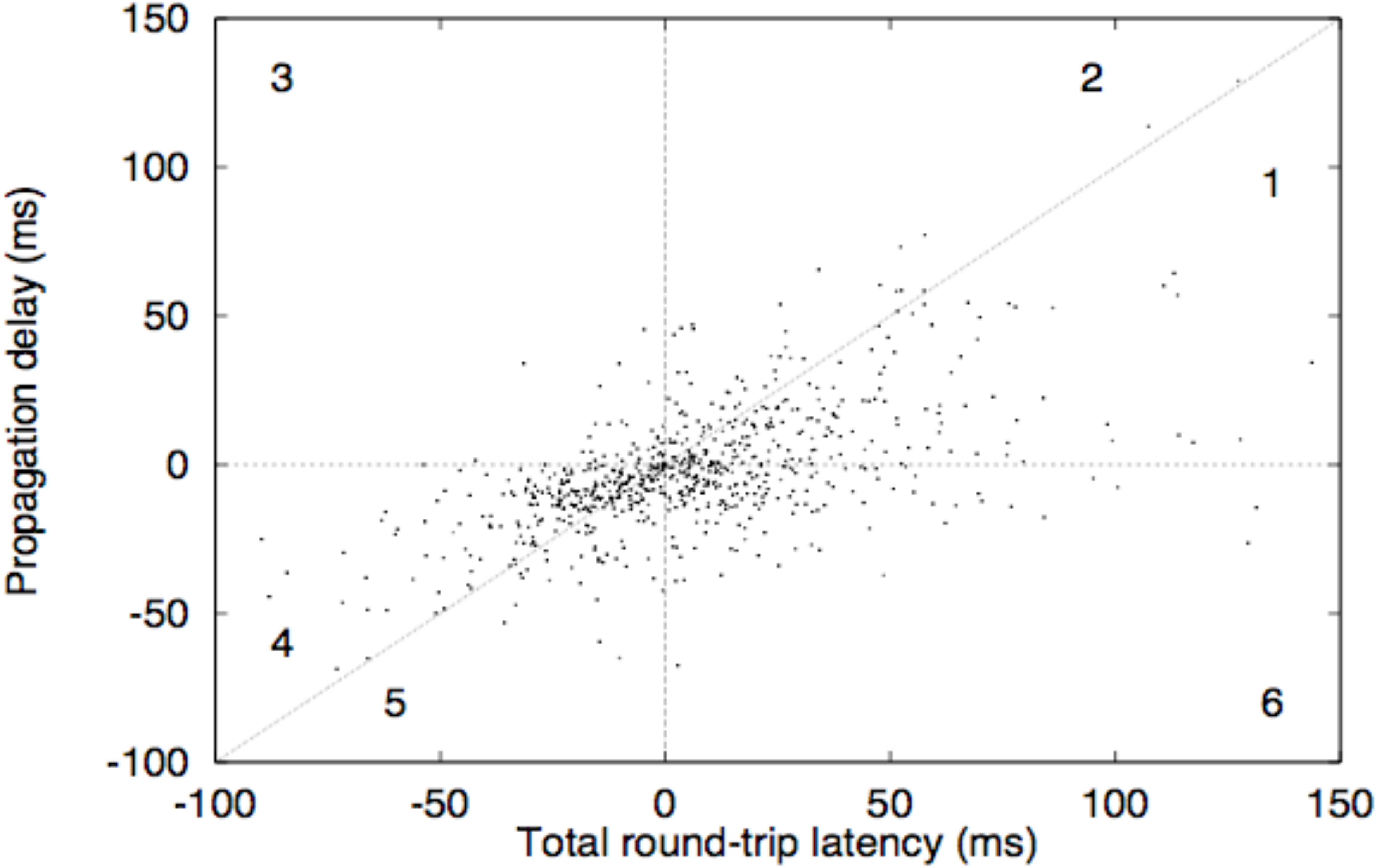


Figure 16: Scatterplot of the portion of the difference in mean round-trip time between the best alternate path and the default path due to the difference in physical propagation delay. Each point represents one pair of hosts. This graph is for the UW3 dataset.

Conclusion:

Can often find better path by routing through another host

Impact:
Inspired overlay networks
(Skype, PPLive, etc.)