End-to-End Routing Behavior in the Internet

Vern Paxson

Network Research Group
Lawrence Berkeley National Laboratory
EECS Division, U.C. Berkeley
vern@ee.lbl.gov

February 15, 1996

Overview:

An experiment to measure *end-to-end* routing behavior: what the users see.

Routing pathologies.

Route stability.

Route symmetry.

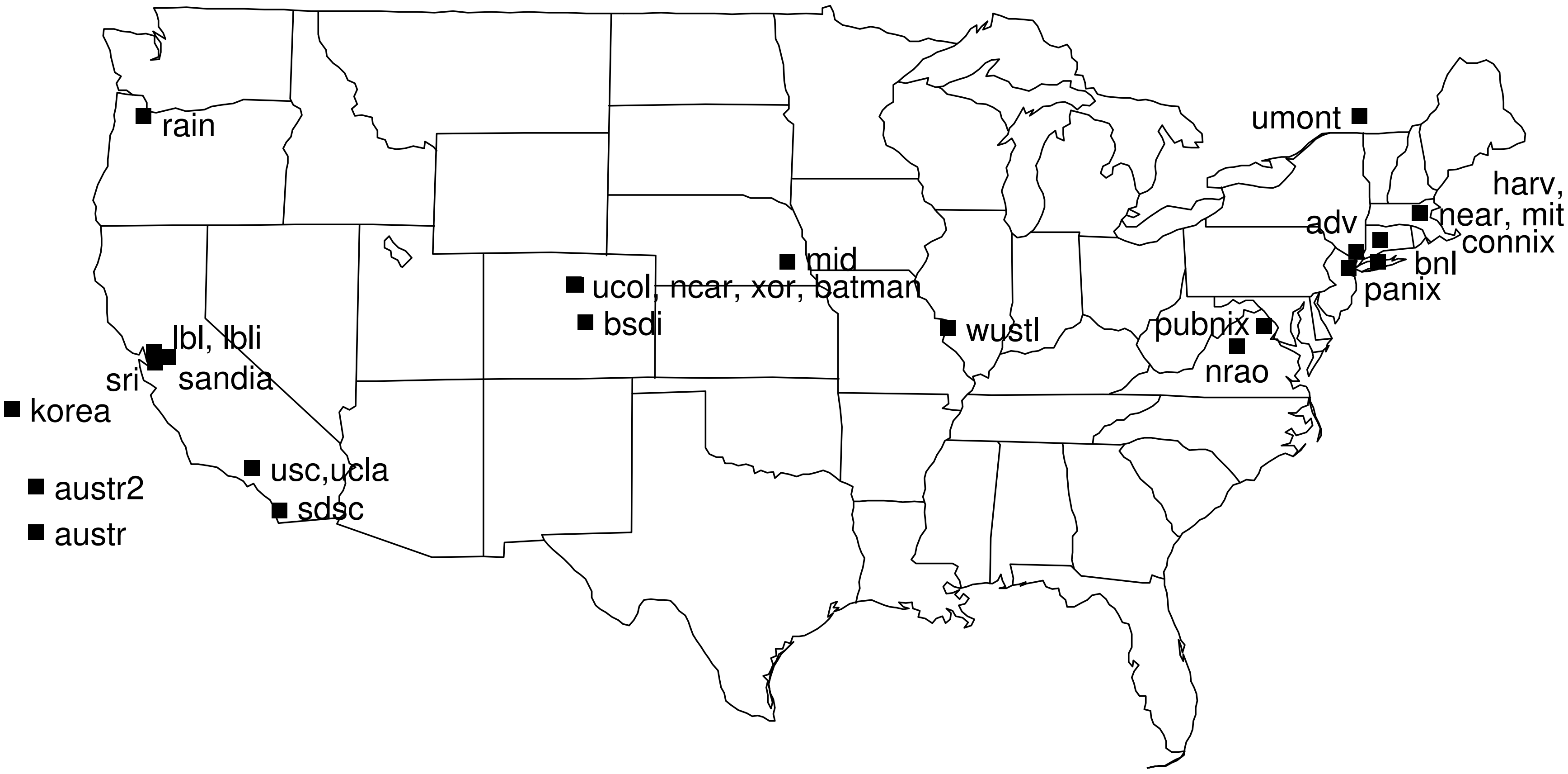
The Network Probe Daemon (NPD):

NPD runs at different sites around the Internet.

It accepts (authenticated) requests to:

- traceroute to a remote host
- run tcpdump/rtcpdump to monitor a probe
- source/sink a probe
- send/reset log
- self-test

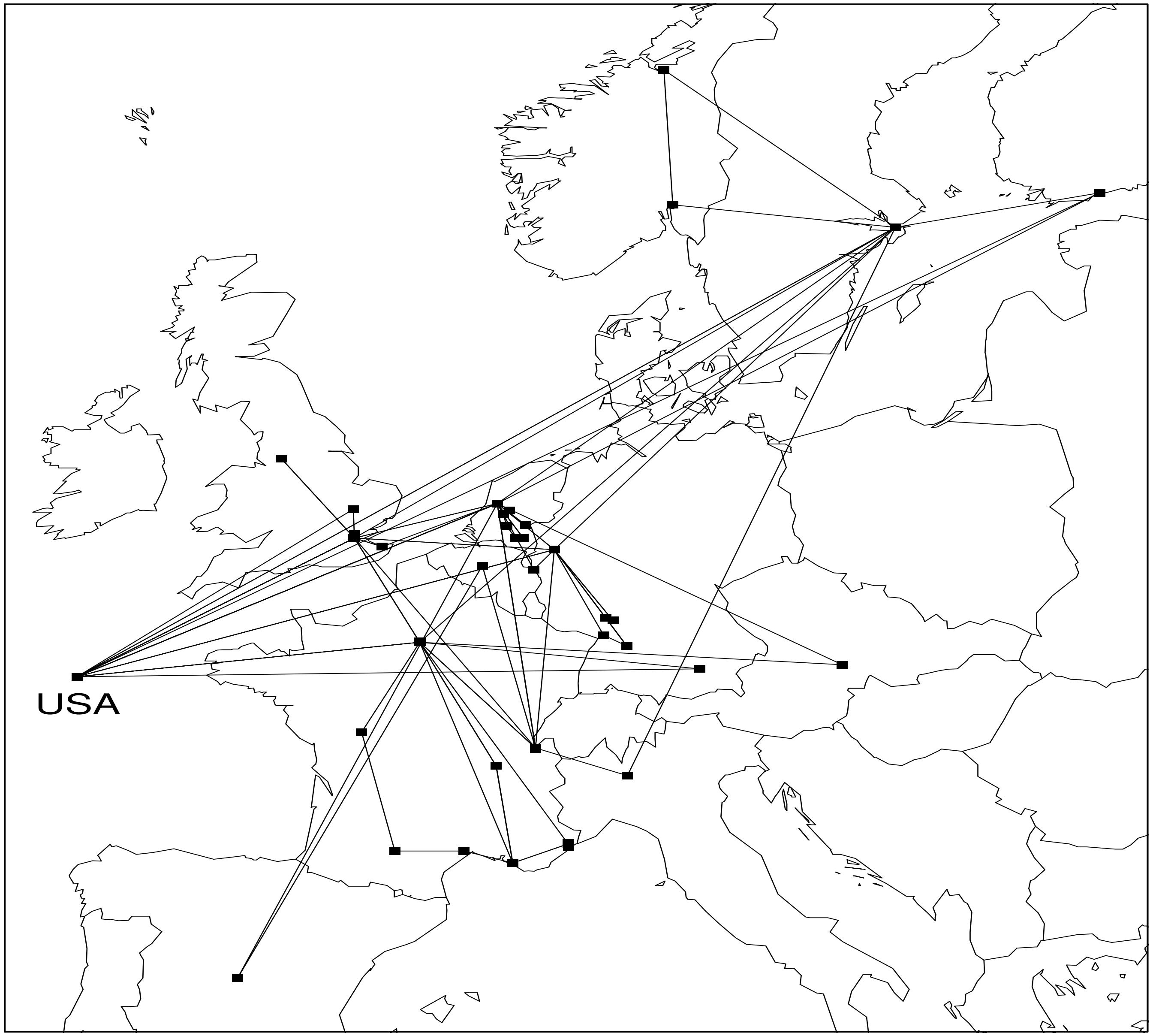
Participating sites — North America



Participating sites — Europe

Links traversed — North America

Links traversed — Europe



The raw data:

$\mathcal{D}_1$: 6,459 `traceroutes` between 27 sites, Nov–Dec 1994.

Each path sampled every 1–2 days.

$\mathcal{D}_2$: 35,109 `traceroutes` between 33 sites, Nov–Dec 1995.

60% of paths sampled w/ mean 2 hours, 40% w/ mean 2.75 days.

85 AS's sampled (about 8% of total active AS's).

Weighted sample: 50+%.

Pathologies — measurement headaches:

Unresponsive routers: rare in $\mathcal{D}_1$, absent in $\mathcal{D}_2$.
(Except weird `adv` / `icm-dante-e0.icp.net` failure.)

Rate-limiting: just in Solaris end-systems and `ucol` router.

Couldn't connect to remote NPD: 5–7%. Introduces a bias towards *underestimation*.

Pathologies — loops:

12 in $\mathcal{D}_1$, 73 in $\mathcal{D}_2$. Some long-lived:

`stuttgart1.belwue.de` $\Leftrightarrow$
`stuttgart4.belwue.de`, 16–32 hr.

`sl-dc-8-h1/0-t3.sprintlink.net` $\Leftrightarrow$
`sl-mae-e-h2/0-t3.sprintlink.net`, $\geq$ 10 hr.

`amsterdam.nl.net` $\Leftrightarrow$ `wgm01.nl.net`, 14–17 hr.

Others: $\leq$ 3 hr and some $\leq$ 10 secs.

Largest loop: 5 routers.

Number of inter-AS loops: 0.

Pathologies — multiple, widespread changes:

```
 1  cs-gw-srl.cs.colorado.edu  3 ms
 2  cu-gw-fddi.colorado.edu  5 ms
 3  ncar-cu.co.westnet.net  13 ms
 4  ml-t3-gw.ucar.edu  11 ms
 5  border2-hssi1-0.denver.mci.net  73 ms
 6  core-fddi-1.denver.mci.net  80 ms
 7  * core2-hssi-2.westorange.mci.net  47 ms
 8  core2-hssi-2.washington.mci.net  58 ms
 9  borderx2-fddi-1.washington.mci.net  73 ms
10  mae-east-plusplus-two.washington.mci.net  60 ms
```

Pathologies — multiple, widespread changes, con't:

```
11  icm-dc-e-f0/0.icp.net  112 ms
12  icm-dc-1-h1/0-t3.icp.net  81 ms
13  icm-dante-e0.icp.net  115 ms  150 ms *
14  * amsterdam1.dante.net  205 ms *
15  nl-s1.dante.bt.net  177 ms
16  nl-f0-0.eurocore.bt.net  172 ms
17  de-s1-1.eurocore.bt.net  206 ms
18  de-f0.dante.bt.net  251 ms
19  * * *
20  * * *
```

Pathologies — multiple, widespread changes, con't:

```
21  * icm-dc-2b-f2/0.icp.net  151 ms
22  icm-dc-1-f0/0.icp.net  97 ms
23  icm-dc-2b-f2/0.icp.net   98 ms
24  icm-dc-1-f0/0.icp.net  109 ms
    umd2-pppl2.es.net  251 ms
25  * mae-east-plusplus-two.washington.mci.net  178
26  pppl2-umd2.es.net  702 ms * *
27  core-hssi-3.sanfrancisco.mci.net  158 ms !H *
    core-fddi-1.denver.mci.net  34 ms !H
```

Pathologies — erroneous routing:

```
 1  mfd-01.rt.connix.net  8 ms
 2  sl-dc-5-s2/0-512k.sprintlink.net  39 ms
 3  sl-dc-6-f0/0.sprintlink.net  39 ms
 4  psi-mae-east-1.psi.net  48 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  * * *
```

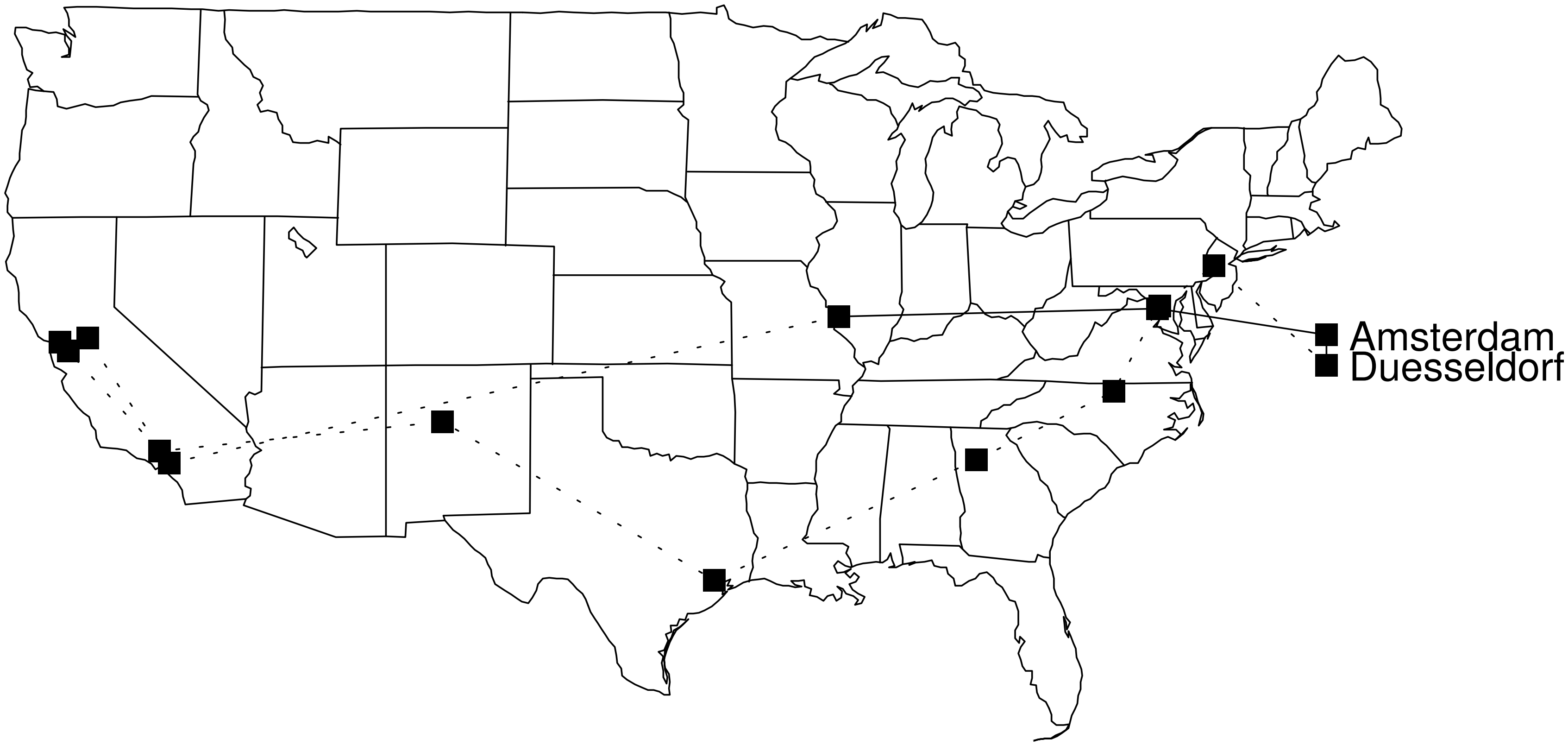
Pathologies — erroneous routing, con't:

```
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
17  cisco101e01.cc.huji.ac.il  2132 ms * *
18  cisco102e13.huji.ac.il  888 ms
19  cisco103e2.gr.huji.ac.il  1657 ms * *
20  * * cisco101e01.cc.huji.ac.il  1349 ms
21-30  * * *
```

Pathologies — duplicate hops:

```
 1  fpls.postech.ac.kr  2 ms  1 ms  1 ms
 2  fddicc.postech.ac.kr  3 ms  2 ms  2 ms
 3  ktrc-postech.hana.nm.kr  30 ms  30 ms  51 ms
 4  gateway.hana.nm.kr  31 ms  31 ms  31 ms
 5  hana.hana.nm.kr  33 ms  44 ms  32 ms
 6  bloodyrouter.hawaii.net  1152 ms  1275 ms  968
 7  bloodyrouter.hawaii.net  744 ms  336 ms  325 ms
 8  arc1.nsn.nasa.gov  384 ms  491 ms  691 ms
 9  jpl6.nsn.nasa.gov  791 ms  772 ms  1082 ms
10  ...
```

Pathologies — flutter:



Problems with large-scale flutter:

Path properties difficult to predict or characterize.

Path can be partially asymmetric.

Out-of-order packets $\Rightarrow$ spurious TCP fast retransmit.

Bodes ill for WAN deflection routing.

None of these a big deal for localized flutter.

Pathologies — “skipping”

```
1 xor-gw.xor.com 0 ms
2 gw1.boulder.co.coop.net 0 ms
3 sl-fw-2-s9-t1.sprintlink.net 30
4 sl-fw-5-f1/0.sprintlink.net 30 ms
5 sl-dc-8-h3/0-t3.sprintlink.net 60 ms
6 icm-dc-1-f0/0.icp.net 1520 ms
  icm-london-1-s1-1984k.icp.net 160 ms
  icm-dc-1-f0/0.icp.net 60 ms
7 icm-london-1-s1-1984k.icp.net 150 ms
8 smds-gw.ulcc.ja.net 140 ms
9 etc.
```

Internet diameter:

Current diameter $>$ 30 hops.

Do hops = kilometers? Not always:

`ncar` $\Rightarrow$ `sdsc`: 3 hops, 1,500 km.

`pubnix` $\Rightarrow$ `bsdi`: 5 hops, 2,000 km.

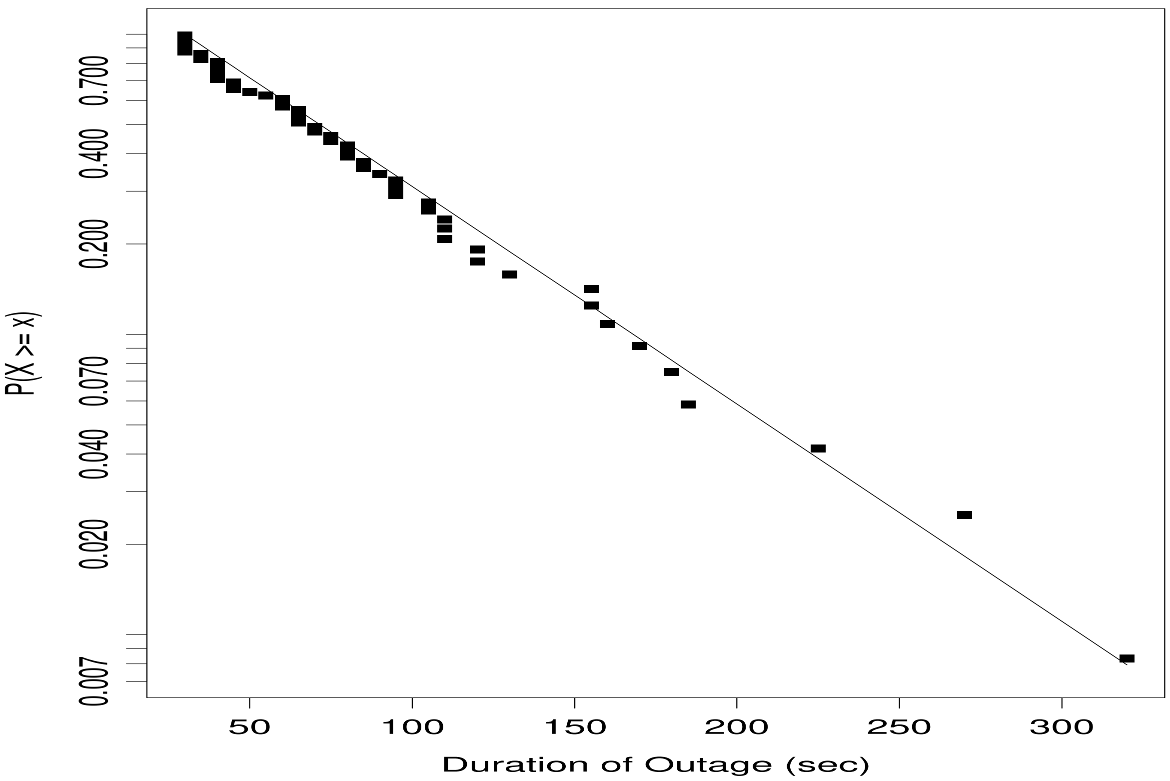
`mit` $\Leftrightarrow$ `harv`: 11 hops, 3 km.

`sri` $\Rightarrow$ `lbl`: up to 17 hops, 50 km.

Temporary outages:

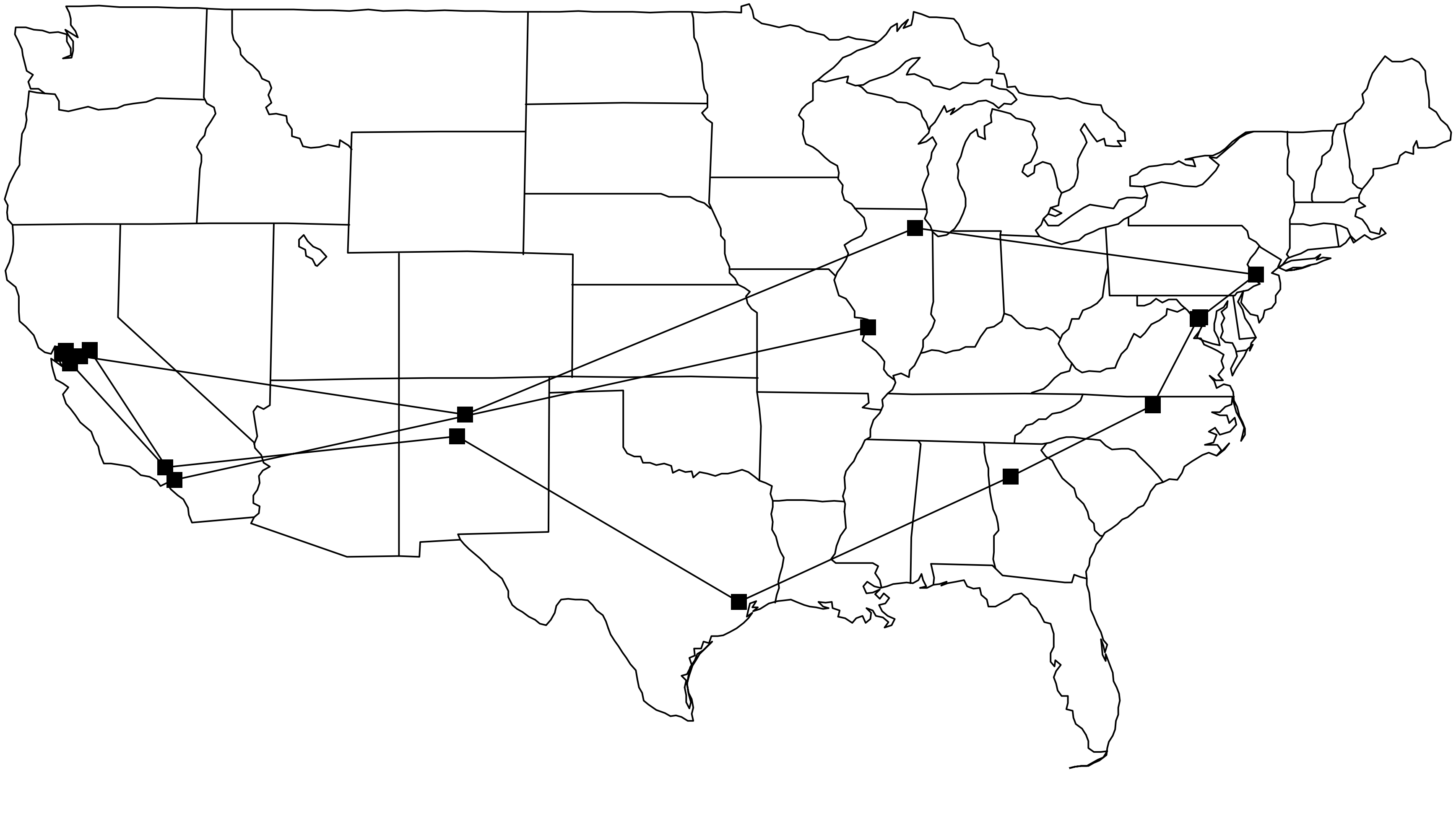
30 secs. + geometric w/ $p = 0.92$ ($\mathcal{D}_1$).

30 secs. + geometric w/ $p = 0.62$ if < 70 secs,
or 70 secs. + geometric w/ $p = 0.92$ if greater ($\mathcal{D}_2$).



Circuitous routing:

`lbl` $\Rightarrow$ `wustl` (ESNET $\Rightarrow$ ANS $\Rightarrow$ SprintLink):



Summary of pathologies:

Pathology	Probability	Trend
Unresponsive routers	0.53% // 0.00%	better
Failure to decrement TTL	0.18% // 0.06%	better
Persistent routing loops	0.13–0.16%	
Temporary routing loops	0.055–0.078%	
Erroneous routing	0.004–0.004%	
Connectivity altered mid-stream	0.16% // 0.44%	worse
Infrastructure failure	0.21% // 0.48%	worse
Temporary outage ≥ 30 secs	0.96% // 2.2%	worse
Total user-visible pathologies	1.5% // 3.4%	worse

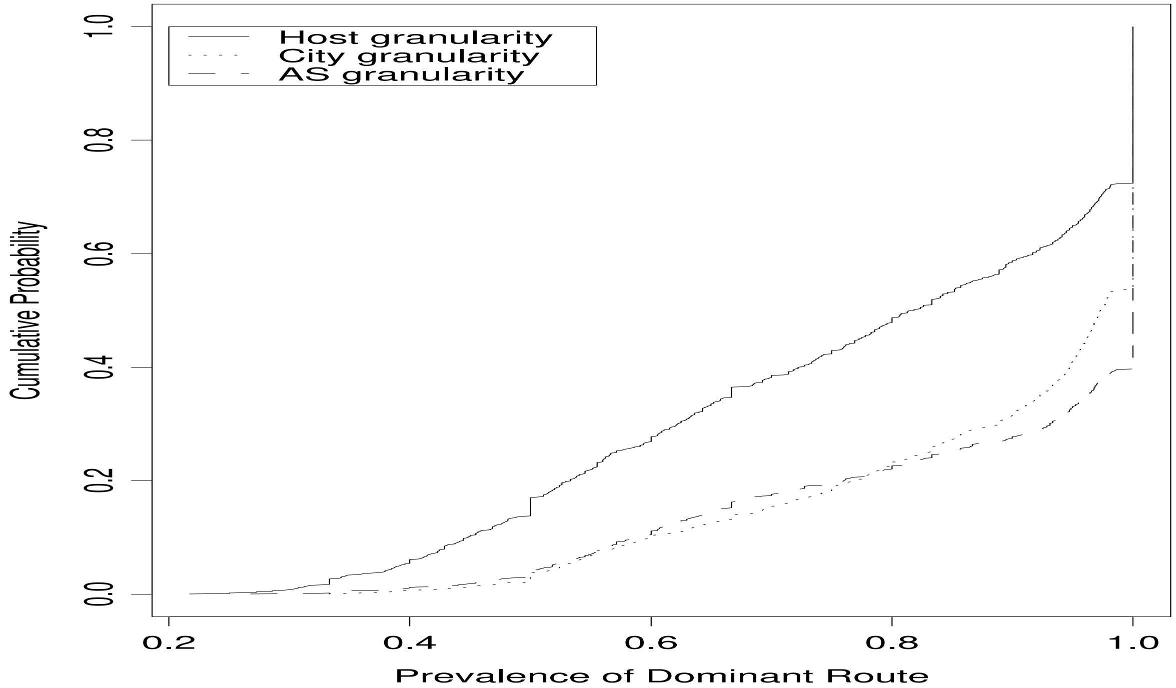
Two types of routing stability:

Prevalence: likelihood of observing a particular route. Easy to assess. Affects utility of caching.

Persistence: how long a route remains unchanged. Hard to assess. Affects utility of storing state in routers; consistency of repeated network measurements.

Route prevalence:

Internet paths are strongly dominated by a single route.

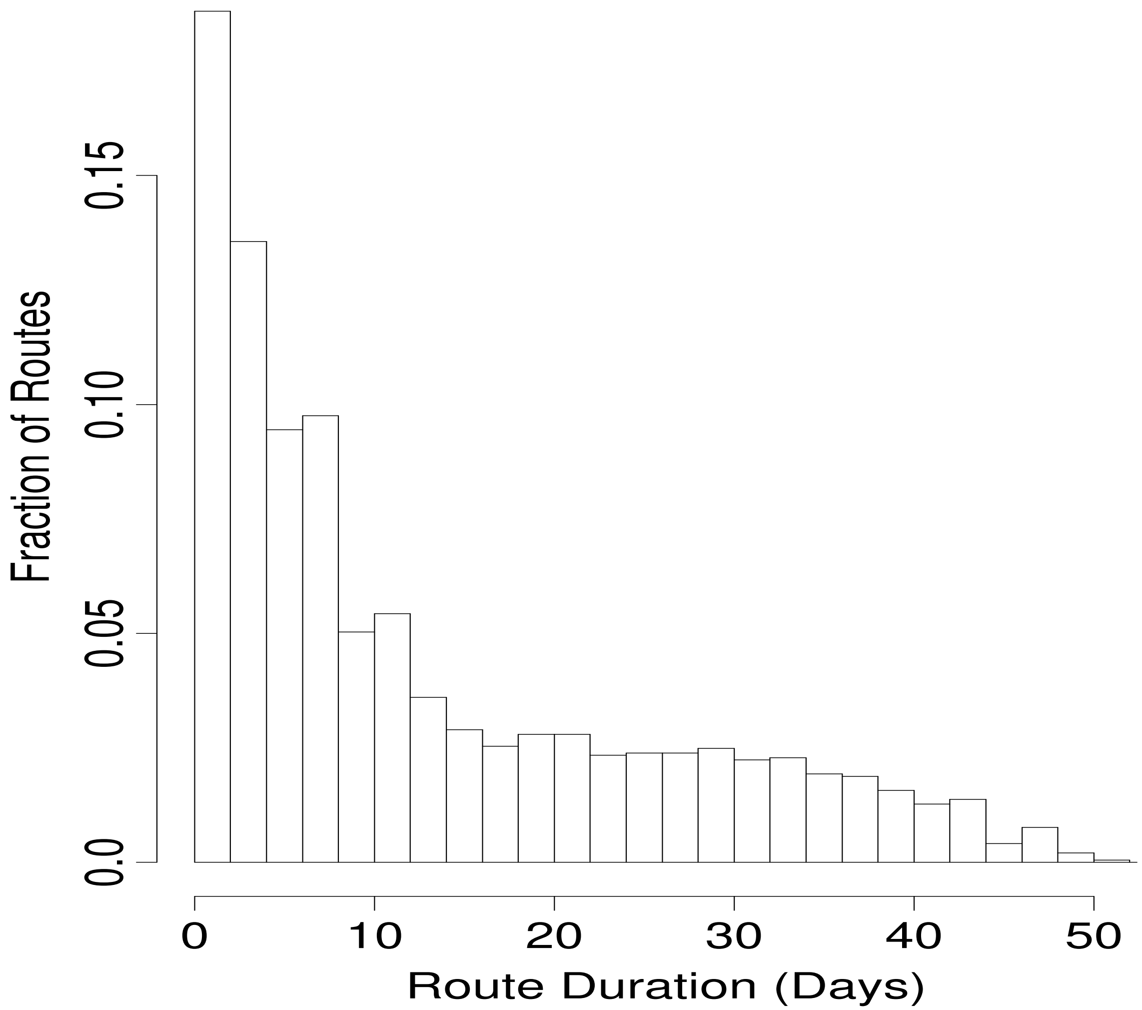


But, as with many aspects of Internet behavior, we also find significant site-to-site variation.

Route persistence:

Time scale	% Paths	Notes
seconds	N/A	Load-balancing “flutter.”
minutes	N/A	“Tightly-coupled” routers.
10's of minutes	9%	Some involved different cities, AS's.
hours	4%	Usually intra-network changes.
6+ hours	19%	Also intra-network changes.
days	68%	Bimodal.

Duration of long-lived routes:



Detecting route changes via TTL:

Granularity	False positives	False negatives	Error rate
host	0%	25%	3%
city	4%	26%	5%
AS path	5%	10%	5%

Summary: it's a decent heuristic, but don't bet the farm on it.

Route symmetry:

Symmetry affects: assumption that one-way latency is RTT/2.

If widespread, presents problems for NTP.

Establishing "anticipatory state" in routers for bidirectional connections.

About 30% of the paths in $\mathcal{D}_1$ visited different cities.

About 50% of the paths in $\mathcal{D}_2$ did so.

30% of the paths in $\mathcal{D}_2$ visited different AS's.

Magnitude of asymmetries:

Most are just a single city or AS.

But 20% of all routes differ in two or more cities in each direction. These can be large: e.g., `ucl` $\Leftrightarrow$ `umann` difference is Amsterdam and Heidelberg vs. Princeton and College Park!

Summary:

Pathologies doubled during 1995, now 1 in 30.

Asymmetries nearly doubled during 1995.

Router software improved during 1995.

Internet paths heavily dominated by a single route.

2/3's of Internet paths are quite stable (days or weeks duration). But the other 1/3 varies over many time scales.

Acknowledgements:

Guy Almes and Bob Camm (`adv`); Jos Alsters (`unij`); Jean-Chrysostome Bolot (`inria`); Hans-Werner Braun, Kim Claffy, and Bilal Chinoy (`sdsc`); Randy Bush (`rain`); Jon Crowcroft and Atanu Ghosh (`ucl`); Peter Danzig and Katia Obraczka (`usc`); Mark Eliot (`sri`); Robert Elz (`austr`); Teus Hagen (`oce`); Steinar Haug and Håvard Eidnes (`sintef1`, `sintef2`); John Hawkinson (`near` and `panix`); TR Hein (`xor`); Tobias Helbig and Werner Sinze (`ustutt`); Paul Hyder (`ncar`); Alden Jackson (`sandia`); Kate Lance (`austr2`); Craig Leres (`lbl`); Kurt Lidl (`pubnix`); Peter Linington, Alan Ibbetson, Peter Collinson,

Acknowledgements, con't:

and Ian Penny (`ukc`); Steve McCanne (`lbli`); John Milburn (`korea`); Walter Mueller (`umann`); Evi Nemeth, Mike Schwartz, Dirk Grunwald, Lynda McGinley (`ucol`, `batman`); Francóis Pinard (`umont`); Jeff Polk and Keith Bostic (`bsdi`); Todd Satogata (`bnl`); Doug Schmidt and Miranda Flory (`wustl`); Sorell Slaymaker and Alan Hannan (`mid`); Don Wells and Dave Brown (`nrao`); Gary Wright (`connix`); John Wroclawski (`mit`); Cliff Young and Brad Karp (`harv`); and Lixia Zhang, Mario Gerla, and Simon Walton (`ucla`).