End-to-End Routing Behavior in the Internet

Vern Paxson
University of California, Berkeley and
Lawrence Berkeley National Laboratory*
vern@ee.lbl.gov

Abstract

The large-scale behavior of routing in the Internet has gone virtually without any formal study, the exception being Chinoy's analysis of the dynamics of Internet routing information [Ch93]. We report on an analysis of 40,000 end-to-end route measurements conducted using repeated "traceroutes" between 37 Internet sites. We analyze the routing behavior for pathological conditions, routing stability, and routing symmetry. For pathologies, we characterize the prevalence of routing loops, erroneous routing, infrastructure failures, and temporary outages. We find that the likelihood of encountering a major routing pathology more than doubled between the end of 1994 and the end of 1995, rising from 1.5% to 3.4%. For routing stability, we define two separate types of stability, "prevalence," meaning the overall likelihood that a particular route is encountered, and "persistence," the likelihood that a route remains unchanged over a long period of time. We find that Internet paths are heavily dominated by a single prevalent route, but that the time periods over which routes persist show wide variation, ranging from seconds up to days. About 2/3's of the Internet paths had routes persisting for either days or weeks. For routing symmetry, we look at the likelihood that a path through the Internet visits at least one different city in the two directions. At the end of 1995, this was the case half the time, and at least one different autonomous system was visited 30% of the time.

1 Introduction

The large-scale behavior of routing in the Internet has gone virtually without any formal study, the exception being Chinoy's analysis of the dynamics of Internet routing information [Ch93]. In this paper we analyze 40,000 end-to-end route measurements conducted using repeated "traceroutes" between 37 Internet sites. The main questions we strive to answer are: What sort of pathologies and failures occur in Internet routing? Do routes remain stable over time or change frequently? Do routes from A to B tend to be symmetric (the same in reverse) as routes from B to A?

*This work was supported by the Director, Office of Energy Research, Scientific Computing Staff, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. A slightly longer version of this paper, with additional figures and typeset with a larger font, is available from `ftp://ftp.ee.lbl.gov/papers/routing.SIGCOMM.ps.Z`.

Our framework for answering these questions is the measurement of a large sample of Internet routes between a number of geographically diverse hosts. We argue that the set of routes is representative of Internet routes in general, and analyze how the routes changed over time to assess how Internet routing in general changes over time.

In § 2 and § 3 we give overviews of related research and how routing works in the Internet. In § 4 we discuss the experimental and statistical methodology for our analysis. In § 5 we give an overview of the participating sites and the raw data. We classify a number of routing pathologies in § 6, including routing loops, rapid routing changes, erroneous routes, infrastructure failures, and temporary outages. We find that the likelihood of encountering a major routing pathology more than doubled between the end of 1994 and the end of 1995, rising from 1.5% to 3.4%.

After removing the pathologies, we analyze the remaining measurements to investigate routing stability (§ 7) and symmetry (§ 8), summarizing our findings in § 9.

2 Related research

The problem of routing traffic in communication networks has been studied for well over twenty years [SS80]. The subject has matured to the point where a number of books have been written thoroughly examining the different issues and solutions [Pe92, St95, Hu95].

A key distinction we will make is that between routing *protocols*, by which we mean mechanisms for disseminating routing information within a network and the particulars of how to use that information to forward traffic, and routing *behavior*, meaning how in practice the routing algorithms perform. This distinction is important because while routing protocols have been heavily studied, routing behavior has not.

The literature contains many studies of routing protocols. In addition to the books cited above, see, for example, discussions of the various ARPANET routing algorithms [MFR78, MRR80, KZ89]; the Exterior Gateway Protocol used in the NSFNET [Ro82] and the Border Gateway Protocol (BGP) that replaced it [RL95, RG95, Tr95a, Tr95b]; the related work by Estrin et al on routing between administrative domains [BE90, ERH92]; Perlman and Varghese's discussion of difficulties in designing routing algorithms [PV88]; Deering and Cheriton's seminal work on multicast routing [DC90]; Perlman's comparison of the popular OSPF and IS-IS protocols [Pe91]; and Baransel et al's survey of routing techniques for very high speed networks [BDG95].

For routing behavior, however, the literature contains considerably fewer studies. Some of these are based on simulation, such as Zaumen and Garcia-Luna Aceves' studies of routing behavior on several different wide-area topologies [ZG-LA92], and Sidhu et al's simulation of OSPF [SFANC93]. In only a few studies do measurements play a significant role: Rekhter and Chinoy's trace-driven simulation of the tradeoffs in using inter-autonomous system routing information to optimize routing within a single autonomous system [RC92]; Chinoy's study of the dynamics of routing information propagated inside the NSFNET infrastructure [Ch93]; and Floyd and Jacobson's analysis of how periodicity in routing messages can lead to global synchronization among the routers [FJ94].

This is not to say that studies of routing protocols ignore routing behavior. But the presentation of routing behavior in the protocol studies is almost always qualitative. Furthermore, of the measurement studies only Chinoy's is devoted to characterizing routing behavior in-the-large.

Chinoy found wide ranges in the dynamics of routing information: For those routers that send updates periodically regardless of whether any connectivity information has changed, the vast majority of the updates contain no new information. Most routing changes occur at the edges of the network and not along its "backbone." Outages during which a network is unreachable from the backbone span a large range of time, from a few minutes to a number of hours. Finally, most networks are nearly quiescent, while a few exhibit frequent connectivity transitions.

Chinoy's study concerns how routing information propagates *inside* the network. It is not obvious, though, how these dynamics translate into the routing dynamics seen by an end user. An area noted by Chinoy as ripe for further study is "the end-to-end dynamics of routing information."

We will use the term *virtual path* to denote the network-level abstraction of a "direct link" between two Internet hosts. For example, when Internet host A wishes to establish a network-level connection to host B, as far as A is concerned the network layer provides it with a link directly to B. We will denote the notion of the virtual path from A to B as $A \Rightarrow B$.

At any given instant in time, the virtual path $A \Rightarrow B$ is realized at the network layer by a single *route*, which is a sequence of Internet routers along which packets sent by A and destined for B are forwarded. Over time, the virtual path $A \Rightarrow B$ may oscillate very rapidly between different routes, or it may be quite stable (c.f. § 7). Chinoy's suggested research area is then: given two hosts A and B at the edges of the network, how does the virtual path $A \Rightarrow B$ behave? This is the question we attempt to answer in our study.

3 Routing in the Internet

For routing purposes, the Internet is partitioned into a disjoint set of *autonomous systems* (AS's) [Ro82]. Originally, an AS was a collection of routers and hosts unified by running a single "interior gateway protocol" (IGP). Over time, the notion has evolved to be essentially synonymous with that of *administrative domain* [HK89], in which the routers and hosts are unified by a single administrative authority, and a set of IGP's. Routing between autonomous systems provides the highest-level of Internet interconnection. RFC 1126 outlines the goals and requirements for inter-AS routing [Li89], and [Re95] gives an overview of how inter-AS routing has evolved.

BGP, currently in its fourth version [RL95, RG95], is now used between all significant AS's [Tr95a]. BGP allows arbitrary interconnection topologies between AS's, and also provides a mechanism for preventing routing loops between AS's (c.f. § 6.1).

The key to whether use of BGP will scale to a very large Internet lies in the *stability* of inter-AS routing [Tr95b]. If routes between AS's vary frequently—a phenomenon termed "flapping" [Do95]—then the BGP routers will spend a great deal of their time updating their routing tables and propagating the routing changes. Daily statistics concerning routing flapping are available from [Me95b].

It is important to note that stable inter-AS routing does *not* guarantee stable end-to-end routing, because AS's are large entities capable of significant internal instabilities.

4 Methodology

In this section we discuss the methodology used in our study: the measurement software; the utility of sampling at exponentially distributed intervals; which aspects of our data are plausibly representative of Internet traffic and which not; how we computed *confidence intervals* for probability estimates; and some problems with our experimental design.

For brevity we assume that the reader is familiar with the workings of the `traceroute` utility for measuring Internet routes ([Ja89]; see [Pa96] for detailed discussion).

4.1 Experimental apparatus

We conducted our experiment by recruiting a number of Internet sites (see Table 1 in § 5) to run a "network probe daemon" (NPD) that provides several measurement services. These NPD's were then periodically contacted by a control program, "npd_control," running on our local workstation, and asked to measure the route to another NPD site using `traceroute` .

For our first set of measurements, termed $\mathcal{D}_1$, we measured each virtual path between two of the NPD sites with a mean interval of 1–2 days. For the second set of measurements, $\mathcal{D}_2$, we made measurements at two different rates: 60% with a mean inter-measurement interval of 2 hours, and 40% with an mean interval of about 2.75 days.

The $\mathcal{D}_1$ interval was chosen so that each NPD would make a `traceroute` measurement on average of once every two hours. As we added NPD sites to the experiment, the rate at which an NPD made measurements to a *particular* remote NPD site decreased, in order to maintain the average load of one measurement per two hours, which led to the range of 1–2 days in the mean measurement interval. Upon analyzing the $\mathcal{D}_1$ data we realized that such a large sampling interval would not allow us to resolve a number of questions concerning routing stability (§ 7). Therefore for $\mathcal{D}_2$ we adopted the strategy of making measurements between pairs of NPD sites in "bursts," with a mean interval of 2 hours between measurements in each burst. We also continued to make lower frequency measurements between pairs of sites in order to gather data to assess routing stability over longer time periods, and arranged the measurements so that 50% would come in bursts and 50% more widely spaced apart. But we also had `traceroute` measurements from a TCP dynamics study we are conducting using the NPD framework (Part II of [Pa96]). These were also made on average two hours apart, so by including them the proportion of burst measurements shifted to 60% bursts, 40% more widely spaced.

The bulk of the $\mathcal{D}_2$ measurements were also *paired*, meaning we would measure the virtual path $A \Rightarrow B$ and then immediately measure the virtual path $B \Rightarrow A$. This enabled us to resolve ambiguities concerning routing symmetry (§ 8), which again we only recognized after having captured and analyzed the $\mathcal{D}_1$ data.

4.2 Exponential sampling

We devised our measurements so that the time intervals between consecutive measurements of the same virtual path were independent and exponentially distributed. Doing so gains two important (and related) properties. The first is that the measurements correspond to *additive random sampling* [BM92]. Such sampling is unbiased because it samples all instantaneous signal values with equal probability. The second important property is that the measurement times form a Poisson process. This means that Wolff's *PASTA principle*—"Poisson Arrivals See Time Averages"—applies to our measurements: asymptotically, the proportion of our measurements that observe a given state is equal to the amount of time that the Internet spends in that state [Wo82]. Two important points regarding Wolff's theorem are (1) the observed process does *not* need to be Markovian; and (2) the Poisson arrivals need not be *homogeneous* [Wo82, § 3].

The only requirement of the PASTA theorem is that the observed process cannot *anticipate* observation arrivals. There is one respect in which our measurements fail this requirement. Even though our observations come exponentially distributed, the network *can* anticipate arrivals as follows: *When the network has lost connectivity between the site running "npd_control" and a site potentially conducting a* `traceroute`*, the network can predict that* no *measurement will occur.* The effect of this anticipation is a tendency to *underestimate* the prevalence of network connectivity problems (see also § 4.5 and § 5.2).

4.3 Which observations are representative?

37 Internet hosts participated in our routing study. This is a miniscule fraction of the estimated 6.6 million Internet hosts as of July, 1995 [Lo95], so clearly behavior we observe that is due to the particular endpoint hosts in our study is not representative. Similarly, the 34 different stub networks to which these hosts belong are also a miniscule fraction of the more than 50,000 known to the NSFNET in April, 1995 [Me95a].

On the other hand, we argue that the *routes* between the 37 hosts are plausibly representative, because they include a non-negligible fraction of the AS's which together comprise the Internet. We expect the different routes within an AS to have similar characteristics (e.g., prevalence of pathologies, or routing stability), because they fall under a common administration, so sampling a significant number of AS's lends representational weight to a set of measurements.

By analyzing a BGP routing table dump obtained from an AS border router, we find in [Pa96] that the Internet presently has about 1,000 active AS's, of which the routes in our study traversed 8%. An important point, however, is that not all AS's are equal—some are much more prominent in Internet routing than others. If we weight each AS by its likelihood of occurring in an AS path, then the AS's sampled by the routes we measured represent about half of the Internet AS's, indicating that our observations are plausibly representative of Internet routing as a whole.

4.4 Confidence intervals

Often in our study we will want to assign some sort of confidence interval to a probability derived from analyzing our data. Suppose that out of a representative sample of n observations we find that a subset of size k exhibit some property $\mathcal{P}$. We might then estimate the unconditional probability p of observing $\mathcal{P}$ as $\hat{p} = k/n$. But the value of $\hat{p}$ is not of much use unless we also have an idea of its possible error. For example, if, out of 2 observations, 1 of them exhibits $\mathcal{P}$, we would not feel too confident declaring that $p \approx \frac{1}{2}$.

To address this problem, we need to associate a *confidence interval* with $\hat{p}$, the interval being a range of values that, with high confidence, includes p. In [Pa96], we develop tight bounds on the interval in which p must lie to be consistent, with confidence c, with observing k independent instances of $\mathcal{P}$ in n measurements. We find that p_l, the lower range of p, is given by:

$$p_l = \frac{\nu_2}{\nu_2 + \nu_1 Q_{F(\nu_1,\nu_2)}(1-c)}$$

where: $\nu_1 = 2(n-k+1)$ and $\nu_2 = 2k$, and $Q_{F(\nu_1,\nu_2)}(1-c)$ is the $1-c$ quantile of the well-known F variance-ratio distribution with parameters ν_1 and ν_2. The upper bound, p_u, has a similar form.

We also look at the problem of *comparing* confidence intervals. Suppose we have two separate datasets, $\mathcal{D}_1$ and $\mathcal{D}_2$, in which we observe k_1 instances of $\mathcal{P}$ out of n_1 independent measurements for $\mathcal{D}_1$, and k_2 out of n_2 for $\mathcal{D}_2$. If we then let c denote the confidence we wish to associate with a finding that the two datasets show a significant difference (i.e., c is the probability that an apparent difference is not simply due to chance), then in [Pa96] we show that we should compute confidence intervals for $\mathcal{D}_1$ and $\mathcal{D}_2$ using $c' = 1 - 2\sqrt{1-c}$. If these intervals do not overlap, then the prevalence of $\mathcal{P}$ in $\mathcal{D}_1$ is significantly different than in $\mathcal{D}_2$, with confidence c.

Throughout our study we use 95% confidence intervals, corresponding to $c = 0.95$ and $c' \approx 0.553$.

4.5 Shortcomings of the experimental design

An understandable criticism of our study is that it does not provide enough analysis of the routing difficulties uncovered, including whether these difficulties are fundamental to routing a large packet-switched internetwork, or whether they could be fixed. There are several reasons for this shortcoming worth noting for those who would undertake a similar study in the future.

The first difficulty is somewhat inherent to end-to-end measurement: while an end-to-end measurement has the great benefit of measuring a quantity of direct interest to network end users, it also has the difficulty of compounding effects at different hops at the network into a single net effect. For example, when a routing loop is observed, a natural question is: what router is responsible for having created this loop? A measurement study made internal to the network, such as [Ch93], can attempt to answer this question because the network's internal state is more visible. But for an end-to-end measurement study such as ours, all that is actually visible is the *fact* that a loop occurs, with little possibility of determining *why*.

One way to determine *why* a problem exists is to ask those running the network. We attempted a great deal of this (see § 10), but this approach does not scale effectively for large numbers of problems.

In retrospect, there are two ways in which our experiment could be considerably improved. The first is that if NPD's could be given a whole batch of measurement requests (rather than just a single request), along with times at which to perform them, then the underestimation of network problems due to our centralized design (§ 4.2) could be eliminated. The second is the use of a tool more sophisticated than `traceroute`: one that could analyze the route

Name	Description
adv	Advanced Network & Services, Armonk, NY
austr	University of Melbourne, Australia
austr2	University of Newcastle, Australia
batman	National Center for Atmospheric Research, Boulder, CO
bnl	Brookhaven National Lab, NY
bsdi	Berkeley Software Design, Colorado Springs, CO
connix	Caravela Software, Middlefield, CT
harv	Harvard University, Cambridge, MA
inria	INRIA, Sophia, France
korea	Pohang Institute of Science and Technology, South Korea
lbl	Lawrence Berkeley Lab, CA
lbli	LBL computer connected via ISDN, CA
mid	MIDnet, Lincoln, NE
mit	Massachusetts Institute of Technology, Cambridge, MA
ncar	National Center for Atmospheric Research, Boulder, CO
near	NEARnet, Cambridge, Massachusetts
nrao	National Radio Astronomy Observatory, Charlottesville, VA
oce	Oce-van der Grinten, Venlo, The Netherlands
panix	Public Access Networks Corporation, New York, NY
pubnix	Pix Technologies Corp., Fairfax, VA
rain	RAINet, Portland, Oregon
sandia	Sandia National Lab, Livermore, CA
sdsc	San Diego Supercomputer Center, CA
sintef1	University of Trondheim, Norway
sintef2	University of Trondheim, Norway
sri	SRI International, Menlo Park, CA
ucl	University College, London, U.K.
ucla	University of California, Los Angeles
ucol	University of Colorado, Boulder
ukc	University of Kent, Canterbury, U.K.
umann	University of Mannheim, Germany
umont	University of Montreal, Canada
unij	University of Nijmegen, The Netherlands
usc	University of Southern California, Los Angeles
ustutt	University of Stuttgart, Germany
wustl	Washington University, St. Louis, MO
xor	XOR Network Engineering, East Boulder, CO

Table 1: Sites participating in the study

measurement in real-time and repeat portions (or all) of the measurement as necessary in order to resolve ambiguities.

5 The Raw Routing Data

5.1 Participating sites

The first routing experiment was conducted from November 8 through December 24, 1994. During this time, we attempted 6,991 `traceroutes` between 27 sites. We refer to this collection of measurements as $\mathcal{D}_1$. The second experiment, $\mathcal{D}_2$, went from November 3 through December 21, 1995. It included 37,097 attempted `traceroutes` between 33 sites. Both datasets are available from the Internet Traffic Archive, `http://town.hall.org/Archives/pub/ITA/`. Table 1 lists the sites participating in our study, giving the abbreviation we will use to refer to the site, a brief description of the site, and its location.

5.2 Measurement failures

In the two experiments, between 5–8% of the `traceroutes` failed outright (i.e., we were unable to contact the remote NPD, execute `traceroute` and retrieve its output). Almost all of the failures were due to an inability of npd_control to contact the remote NPD.

For our analysis, the effect of these contact failures will lead to a bias towards *underestimating* Internet connectivity failures, because sometimes the failure to contact the remote daemon will result in losing an opportunity to observe a lack of connectivity between that site and another remote site (§ 4.2).

When conducting the $\mathcal{D}_2$ measurements, however, we somewhat corrected for this underestimation by *pairing* each measurement of the virtual path $A \Rightarrow B$ with a measurement of the virtual path $B \Rightarrow A$, increasing the likelihood of observing such failures. In only 5% of the $\mathcal{D}_2$ measurement failures was npd_control also unable to contact the other host of the measurement pair.

6 Routing pathologies

We begin our analysis by classifying occurrences of routing pathologies—those routes that exhibited either clear, sub-standard performance, or out-and-out broken behavior.

6.1 Routing loops

In this section we discuss the pathology of a routing *loop*. For our discussion we distinguish between three types of loops: a *forwarding* loop, in which packets forwarded by a router eventually return to the router; an *information* loop, in which a router acts on connectivity information derived from information it itself provided earlier; and a `traceroute` loop, in which a `traceroute` measurement reports the same sequence of routers multiple times. For our study, all we can observe directly are `traceroute` loops, and it is possible for a `traceroute` loop to reflect *not* a forwarding loop but instead an upstream routing change that happens to add enough upstream hops that the `traceroute` observes the same sequence of routers as previously. Because of this potential ambiguity, we require a `traceroute` measurement to show the same sequence of routers at least *three* times in order to be assured that the observation is of a forwarding loop.

In general, routing algorithms are designed to avoid forwarding loops, provided all of the routers in the network share a consistent view of the present connectivity. Thus, loops are apt to form when the network experiences a change in connectivity and that change is not immediately propagated to all of the routers [Hu95]. One hopes that forwarding loops resolve themselves quickly, as they represent a complete connectivity failure.

While some researchers have downplayed the significance of temporary forwarding loops [MRR80], others have noted that loops can rapidly lead to congestion as a router is flooded with multiple copies of each packet it forwards [ZG-LA92], and minimizing loops is a major Internet design goal [Li89]. To this end, BGP is designed to never allow the creation of inter-AS forwarding loops, which it accomplishes by tagging all routing information with the AS path over which it has traversed.[1]

Persistent routing loops. For our analysis, we considered any `traceroute` showing a loop unresolved by end of the `traceroute` as a "persistent loop." 10 `traceroutes` in $\mathcal{D}_1$ exhibited persistent routing loops. See [Pa96] for details.

In $\mathcal{D}_2$, 50 `traceroutes` showed persistent loops. Due to $\mathcal{D}_2$'s higher sampling frequency, for some of these loops we can place upper bounds on how long they persisted, by looking for surrounding measurements between the same hosts that do not show the

[1] This technique is based on the observation that forwarding loops occur only in the wake of a routing information loop.

loop. In addition, sometimes the surrounding measurements *do* show the loop, allowing us to assign lower bounds, too.

Source	Dest.	Date	#	Location	Duration
inria	adv	Nov. 6	1	Washington	?
inria	near	Nov. 11	1	Washington	$\leq$ 3 hr
wustl	inria	Nov. 24	1	Washington	?
inria	pubnix	Nov. 12	1	Washington	?
inria	austr2	Nov. 15	1	Washington	?
sintef1	adv	Nov. 12	1	Washington	?
pubnix	sintef1	Nov. 8	1	Anaheim	?
ustutt	ucl	Nov. 11	16	Stuttgart	16–32 hr
connix	bsdi	Nov. 14	1	MAE-East	$\geq$ 10 hr
ustutt	austr	Nov. 14	1	same loop	
pubnix	sintef1	Nov. 14	1	Washington	$\leq$ 5.5 hr
austr	nrao	Nov. 15	1	College Park	?
many	oce	Nov. 23	12	Amsterdam	14–17 hr
ucol	ustutt	Nov. 24	1	San Francisco	?
ucol	inria	Nov. 27	1	Paris	$\leq$ 14 hr
mid	bsdi	Nov. 28	1	Washington	$\leq$ 3 hr
mid	austr	Dec. 6	1	Chicago	$\leq$ 3 hr
mit	wustl	Dec. 10	1	St. Louis	?
umann	nrao	Dec. 13	1	Heidelberg	?
ucl	mit	Dec. 14	1	Cambridge	$\leq$ 3 hr
near	ucla	Dec. 16	1	Los Angeles	?
sri	near	Dec. 17	1*	Palo Alto	?
near	sri	same	1*	San Francisco	?
bsdi	sintef1	Dec. 21	1	NJ, London	$\leq$ 10 hr

Table 2: Persistent routing loops in $\mathcal{D}_2$

Table 2 summarizes the loops seen in $\mathcal{D}_2$. The first two columns give the source and destination of the `traceroute`, the next column the date, the fourth column the number of consecutive `traceroutes` that encountered the loop, and the fifth column the location. Note that only one of the loops spanned multiple cities (and multiple continents!), the last in the table. The final column gives the bounds we were able to assess for the duration of the loop. Loops for which we were unable to assign plausible bounds are marked "?".

The loop durations fall into two modes, those definitely under 3 hours (and possibly quite shorter), and those of more than half a day. The presence of persistent loops of durations on the order of hours is quite surprising, and suggests a lack of good tools for diagnosing network problems.

We also note a tendency for persistent loops to come in clusters. Geographically, loops occurred much more often in the Washington D.C. area, probably because the very high degree of interchange between different network service providers in that area offers ample opportunity for introducing inconsistencies.

Loops involving separate pairs of routers also are clustered in time. The `pubnix` $\Rightarrow$ `sintef1` loop, involving two AlterNet routers sited in Washington D.C., was measured at the same time as the `connix` $\Rightarrow$ `bsdi` and `ustutt` $\Rightarrow$ `austr` observations of a SprintLink loop, at nearby MAE-East. The `sri` $\Rightarrow$ `near` and `near` $\Rightarrow$ `sri` loop observations were paired measurements. They do *not* observe the same loop, but rather two separate loops between closely related routers. Thus it appears that the inconsistencies that lead to long-lived routing loops are not confined to a single pair of routers but also affect nearby routers, tending to introduce loops into their tables too. This clustering makes sense because topologically close routers will often quickly share routing information, and hence if one router's view is inconsistent, the view of the nearby ones is likely to be so, too. The clustering suggests that an observation of a persistent forwarding loop likely reflects an outage of larger scope than just the observed set of looping routers.

Temporary routing loops. We define a temporary loop as one that resolved during the `traceroute`. In $\mathcal{D}_1$ we observed only two temporary loops, but in $\mathcal{D}_2$ we found 23. These are detailed in [Pa96]. Here, we limit the discussion to an interesting property we often found associated with these loops, namely widespread connectivity or routing changes. For example, in a `traceroute` from `rain` to `inria`, we observed a forty second outage; followed by a loop between five MCINET routers sited at Washington, D.C.; followed by a loss of connectivity all the way back to the `rain` border router; followed by connectivity regained all the way to `inria`. It is these middle two events that are surprising, that a loop in Washington resolved into a connectivity outage between Portland and Seattle.

Most likely these widespread changes reflect the "ripple effects" of a single routing transition (a link going down), as a transient connectivity outage propagates through the Internet. This conjecture could be further assessed by an analysis of BGP routing transition statistics, such as those available from [Me95b].

Location of routing loops. We analyzed the looping routers to see if any of the loops involved more than one AS. As mentioned above, the design of BGP in theory prevents any inter-AS forwarding loops, by preventing any looping of routing information. We found that *all* of the $\mathcal{D}_1$ and $\mathcal{D}_2$ routing loops were confined to a single AS, providing solid evidence that BGP route loop suppression works well in practice.

6.2 Erroneous routing

In $\mathcal{D}_1$ we found one example of *erroneous* routing, where the packets clearly took the wrong path. This involved a `connix` $\Rightarrow$ `ucl` route in which the trans-Atlantic hop was not to London but instead to Rehovot, Israel! While we did not observe any erroneous routing in $\mathcal{D}_2$, there remains a security lesson to be considered: one really cannot make any safe assumptions about where one's packets might travel on the Internet.

6.3 Connectivity altered mid-stream

In 10 of the $\mathcal{D}_1$ traces we observed routing connectivity reported earlier in the `traceroute` later lost or altered, indicating we were catching a routing failure as it happened. See [Pa96] for examples. Some of these changes were accompanied by outages, in which presumably the intermediary routers were rearranging their views of the current topology, and dropping many packets in the interim because they did not know how to forward them. We found that the distribution of recovery times from routing problems is at least bimodal—some recoveries occur quite quickly, on the time scale of congestion delays (100's of msec to seconds), while others take on the order of a minute to resolve. The latter type of recovery presents significant difficulties for time-sensitive applications that assume outages are short-lived.

In contrast with the rarity of connectivity changes in $\mathcal{D}_1$(10 total), in $\mathcal{D}_2$ we observed 155 instances of a change, a fact we comment upon further in § 6.10.

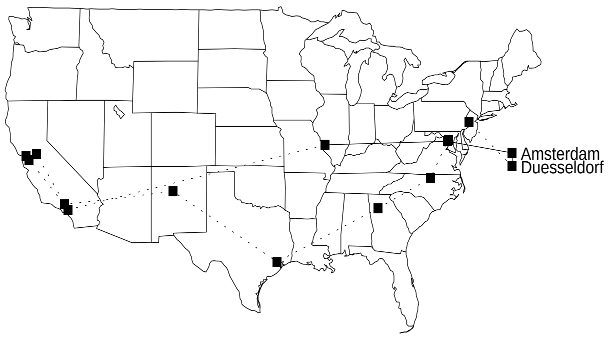

Figure 1: Routes taken by alternating packets from wustl (St. Louis, Missouri) to umann (Mannheim, Germany), due to fluttering

6.4 Fluttering

We use the term "fluttering" to refer to rapidly-oscillating routing. Figure 1 dramatically illustrates the possible effects of fluttering. Here, the wustl border router splits it load between two STARnet routers in St. Louis, one of which sends all of its packets to Washington, D.C. (solid; 17 hops to umann), and the other to Anaheim (dotted line; 29 hops). Thus, every other packet bound for umann travels via a different coast! While load splitting is explicitly allowed in [Ba95, p.79], that document also cautions that there are situations for which it is inappropriate. We argue below that this is one of those situations.

In addition to the wustl fluttering (which occurred in every wustl originated traceroute in $\mathcal{D}_1$, except when the Anaheim link went down), we also found fluttering at a ucol border router. Here, though, the two split paths immediately rejoined, so the split's effects were completely localized. In $\mathcal{D}_2$, however, we saw very little fluttering—see [Pa96] for details and additional discussion.

While fluttering can provide benefits as a way to balance load in a network, it also creates a number of problems for different networking applications. First, a fluttering network path presents the difficulties that arise from *unstable* network paths (§ 7.1). Second, if the fluttering only occurs in one direction, then the path suffers from the problems of *asymmetry* (§ 8.1). Third, constructing reliable estimates of the path characteristics, such as round-trip time and available bandwidth, becomes potentially very difficult, since in fact there may be *two* different sets of values to estimate. Finally, when the two routes have different propagation times, then TCP packets arriving at the destination out of order can lead to spurious "fast retransmissions" [St94] by generating duplicate acknowledgements, wasting bandwidth.

These problems all argue for eliminating large-scale fluttering whenever possible. On the other hand, when the effects of the flutter are confined, as for ucol, or invisible at the network layer (such as split-routing used at the link layer, which would not show up at all in our study), then these problems are all ameliorated. Furthermore, if fluttering is done on a coarser granularity than per packet (say, per TCP connection), then the effects are also lessened.

Finally, we note that "deflection" routing schemes that forward packets along sub-optimal routes to avoid the need to buffer packets at routers, and/or to simplify routing decisions [BDG95], have virtually the *same* characteristics as fluttering paths. In particular, deploying such schemes in wide-area networks could lead to grievous difficulties unless the schemes include mechanisms for tightly controlling the scope of the route differences.

6.5 Infrastructure failures

In addition to traceroute failures due to persistent routing loops and erroneous routing, 125 of the $\mathcal{D}_1$ traceroutes and 617 of the $\mathcal{D}_2$ traceroutes failed to reach the destination host for other reasons. We analyze these failures in detail in [Pa96]. Here, we confine ourselves to "infrastructure failures," in which a route terminates in the middle of the network.

We observed a total of 13 infrastructure failures out of 6,459 $\mathcal{D}_1$ observations, corresponding to an Internet infrastructure availability of 99.7–99.9%, while in $\mathcal{D}_2$ this dropped to 99.4–99.6%. We must bear in mind, however, that these numbers will be somewhat skewed by times when the infrastructure failure also prevented us from making any measurement (§ 5.2), so these availability figures are overestimates.

6.6 Unreachable due to too many hops

By default, traceroute probes up to 30 hops of the route between two hosts. This length sufficed for all of the $\mathcal{D}_1$ measurements, and all but 6 of the $\mathcal{D}_2$ measurements. The fact that it failed occasionally in $\mathcal{D}_2$, however, indicates that the operational diameter of the Internet has grown beyond 30 hops, and argues for using large initial TTL values when a host originates an IP datagram.

It is sometimes assumed that the hop count of a route equates to its geographical distance. While this is roughly the case, we noticed some remarkable exceptions. For example, we observed a 1,500 km end-to-end route of only 3 hops, and a 2,000 km route of 5 hops. We also found that the route between mit and harv (about 3 km apart), was consistently 11 hops in both directions. See [Pa96] for details.

6.7 Temporary outages

The final pathology we discuss here is temporary network outages. When a sequence of consecutive traceroute probes are lost, the most likely cause is either a temporary loss of network connectivity, or very heavy congestion lasting 10's of seconds. For each traceroute, we examined its longest period of consecutive probe losses (other than consecutive losses at the end of a traceroute when, for example, the endpoint was unreachable). The resulting distribution of the number of probes lost appears trimodal. In $\mathcal{D}_1$ ($\mathcal{D}_2$), about 55% (43%) of the traceroutes had no losses, 44% (55%) had between 1 and 5 losses, and 0.96% (2.2%) had 6 or more losses.

Of these latter (six or more losses, $\geq$ 30 sec outage), the distribution of the number of probes lost in the $\mathcal{D}_1$ data is quite close to geometric, with $p = 0.92$ that a probe beyond the 6th is dropped.

In the $\mathcal{D}_2$ data, however, we find that the geometric tail with $p = 0.92$ is present only for outages more than 75 seconds long. For outages between 30 and 70 seconds, the duration still exhibits a strong geometric distribution, but with $p = 0.62$, suggesting two different recovery mechanisms. See [Pa96] for additional discussion. We do not have a plausible explanation for the difference, nor for why the distribution is geometric.

6.8 Time-of-day patterns

We analyzed the two most prevalent pathologies in $\mathcal{D}_2$ for time-of-day patterns, to determine whether they are correlated with the known patterns of heavy traffic levels during daytime hours and lower levels during the evening and early morning off-hours. To do so, we must first associate a time-of-day with a `traceroute` measurement that might span multiple time zones or even continents. We did so by assigning to each measurement the mean of the time-of-day at its source and destination hosts. For example, the time zone of Berkeley, California is three hours behind that of Cambridge, Massachusetts. For a `traceroute` from `mit` to `lbl`, initiated at 09:00 local time in Cambridge, we would assign a local time of 07:30, since the `traceroute` occurred at 06:00 local time in California.

The first question to study is whether the measurements themselves show a time-of-day pattern. In principle, they should not, because the exponential sampling (§ 4.2) is done without regard to the local time, so measurements should occur throughout the day with equal likelihood. However, as discussed in § 4.2, our methodology was flawed in the sense that no measurements were made when our centralized measurement process was unable to contact a remote NPD. Thus we would expect to find a bias in the time-of-day of the measurements towards times of higher connectivity.

Indeed, we find such an effect. By binning each measurement's time-of-day into one of the day's 24 hours, we constructed a histogram of which hours had the most measurements and which the least. We found that the most (4.5%) occurred during the 00:00–01:00 hour, and the least (3.8%) during the 13:00-14:00 hour, with clear correlation between better connectivity and the evening and early morning hours. This finding accords with the widely recognized phenomenon that congestion peaks during working hours, and hence, one might expect, so do connectivity outages. The spread across the course of the day is not too great, however, with the low hour accounting for only 15% fewer connections than the high hour.

The most prevalent pathology was a temporary outage lasting at least 30 seconds (§ 6.7). We would expect these outages to be strongly correlated with the time-of-day congestion patterns. Indeed, this is the case. In $\mathcal{D}_2$, the fewest temporary outages (0.4%) occurred during the 01:00-02:00 hour, while the most (8.0%) occurred during the 15:00-16:00 hour, with the pattern closely following the daily congestion pattern.

The other pathology we analyzed was that of an infrastructure failure (§ 6.5). Here, we again have the peak occurring the 15:00-16:00 hour (9.3%), but the minimum actually occurred during the 09:00-10:00 hour (1.2%). Furthermore, the second highest peak (7.6%) occurred during the 06:00-07:00 hour. We speculate that this pattern might reflect the network operators favoring early morning (before peak hours) for making configuration changes and repairs. Once finished, these then hold the network stable until the late afternoon hours, when congestion hits its peak.

6.9 Representative pathologies

In § 4.3 we argued that our measurements *in general* are plausibly representative. An important question, though, is whether the *pathologies* are likewise representative. It could be that our collection of sites happened to include an atypical AS responsible for much more than its representative share of pathologies. For example, if the regional network associated with one of the sites was more prone to looping than most AS's, then our measurements might observe loops much more often than the frequency by which they occur in the general Internet.

It often proves difficult to assign responsibility for a pathology to a particular AS, in part due to the "serial" nature of `traceroute` (§ 4.5): a pathology observed in a `traceroute` measurement as occurring at hop h might in fact be due to a router upstream to hop h that has changed the route, or a router downstream from h that has propagated inconsistent routing information upstream to h. Nevertheless, we attempted to assess the representativeness of the pathologies as follows. For the most common pathology, a temporary outage of 30 or more seconds (§ 6.7), we assigned responsibility for the outage to the router in the `traceroute` measurement directly upstream from the first completely missing hop, as the link between this router and the missing hop is the most likely candidate for subsequent missing packets. We then tallied for each AS the number of its routers held culpable for outages.

The top three AS's accounted for nearly half of all of the temporary outages. They were AS-3561 (MCI-RESTON), 25%; AS-1800 (ICM-Atlantic; the transcontinental link between North America and Europe, operated by Sprint), 16%; and AS-1239 (Sprintlink), 6%. These three also correspond to the top three AS's by "weight" (§ 4.3), indicating that our observations of the pathology are not suffering from skew due to an atypical AS.

Pathology	Probability	Trend	Notes
Persistent loops	0.13–0.16%		Some lasted hours.
Temporary loops	0.055–0.078%		
Erroneous routing	0.004–0.004%		No instances in $\mathcal{D}_2$.
Mid-stream change	0.16% // 0.44%	worse	Suggests rapidly varying routes.
Infrastructure failure	0.21% // 0.48%	worse	No dominant link.
Outage $\geq$ 30 secs	0.96% // 2.2%	worse	Duration exponent. distributed.
Total pathologies	1.5% // 3.4%	worse	

Table 3: Summary of representative routing pathologies

6.10 Summary of pathologies

Table 3 summarizes the routing pathologies. The second column gives the probability of observing the pathology, in two forms. A range indicates that the proportion of observations in $\mathcal{D}_1$ was consistent with the proportion in $\mathcal{D}_2$ (using the methodology outlined in § 4.4). The range reflects the values consistent with both datasets. Two probabilities separated by "//" indicates that the proportion of $\mathcal{D}_1$ observations was *inconsistent* with the proportion of $\mathcal{D}_2$ observations. The first probability applies to $\mathcal{D}_1$, and reflects the state of the Internet at the end of 1994, and the second to $\mathcal{D}_2$, reflecting the state at the end of 1995.

For those pathologies with inconsistent probabilities, the third column assesses the trend during the year separating the $\mathcal{D}_1$ and $\mathcal{D}_2$ measurements. *None of the pathologies improved!*, and *a number became significantly worse*.

The final row summarizes the total probability of observing a pathology. *During 1995, the likelihood of a user encountering a serious end-to-end routing problem more than doubled, and is now 1 in 30.* The most prevalent of these problems is an outage lasting more than 30 seconds.

This finding should concern anyone interested in the long-term stability of the Internet. While it is always dangerous to infer a trend from only two points, clearly if the pattern is indeed a trend, then network service will degrade to unacceptable levels. An argument that it might not be a trend is that 1995 was an atypical year for Internet stability, due to the transition from the NSFNET backbone to the commerically-operated backbone. An argument that it is a trend, however, comes from recent data indicating increasing inter-AS routing instability during the second quarter of 1996 [La96].

7 End-to-end routing stability

One key property we would like to know about an end-to-end Internet route is its *stability*: do routes change often, or are they stable over time? In this section we analyze the routing measurements to address this question. We begin by discussing the impact of routing stability on different aspects of networking. We then present two different notions of routing stability, "prevalence" and "persistence," and show that they can be independent. It turns out that "prevalence" is quite easy to assess from our measurements, and "persistence" quite difficult. In § 7.4 we characterize the prevalence of Internet routes, and then in § 7.5 we tackle the problem of assessing persistence.

7.1 Importance of routing stability

One of the goals of the Internet architecture is that large-scale routing changes (i.e., those involving different autonomous systems) rarely occur [Li89]. There are a number of aspects of networking affected by routing stability: the degree to which the properties of network paths are *predictable*; the degree to which a connection can *learn* about network conditions from past observations; the degree to which real-time protocols must be prepared to recreate or migrate state stored in the routers [DB95, FBZ94, ZDESZ93, BCS94]; and the degree to which network studies based on repeated measurements of network paths ([CPB93, Bo93, SAGJ93, Mu94]) can assume that the measurements are indeed observing the same path.

7.2 Two definitions of stability

There are two distinct views of routing stability. The first is: "Given that I observed route r at the present, how likely am I to observe r again in the future?" We refer to this notion as *prevalence*, and equate it with the probability of observing a given route. Prevalence has implications for overall network predictability, and the ability to learn from past observations (c.f. § 7.1).

A second view of stability is: "Given that I observed route r at time t, how long before that route is likely to have changed?" We refer to this notion as *persistence*. It has implications for how to effectively manage router state, and for network studies based on repeated path measurements.

Intuitively, we might expect these two notions to be coupled. Consider, for example, a sequence of routing observations made every T units of time. If the routes we observe are:

$$R_1, R_1, R_1, R_1, R_1, R_1, R_1, R_1, R_1, R_1, R_1, R_2, R_1, R_1, R_1 \ldots$$

then clearly route R_1 is much more prevalent than route R_2. We might also conclude that route R_1 is persistent, because we observe it so frequently; but this is not at all necessarily the case. For example, suppose T is one day. If the mean duration of R_1 is 10 days, and that of R_2 is one day, then this sequence of observations is quite plausible, and we would be correct in concluding that R_1 is *persistent and prevalent*. Furthermore, depending on our concern, we might also deem that R_2 is persistent, since on average it lasts for a full day. In that case, R_2 is *persistent but not prevalent.*

But suppose instead that the mean duration of R_1 is 10 seconds and the mean duration of R_2 is 1 second. If, for example, the alternations between them occur as a semi-Markov process, then the proportion of time spent in state R_1 is $\frac{10}{11}$ [Ro83], again reflecting that R_1 is prevalent. Similarly, the proportion of time spent in state R_2 is $\frac{1}{11}$. Given these proportions, the sequence of observations is *still plausible*, even though each observation of R_1 is actually of a separate instance of the route. In this case, R_1 is *prevalent but not persistent*, and R_2 is *neither prevalent nor persistent*.

7.3 Reducing the data

We confine our analysis to the $\mathcal{D}_2$ measurements, as these were made at a wide range of intervals (60% with mean 2 hrs and 40% with mean 2.75 days), which allows us to assess stability over many time scales, and to tackle the "persistence ambiguity" outlined above. Of the 35,109 $\mathcal{D}_2$ measurements, we omitted those exhibiting pathologies (because they reflect difficulties distinct from routing instabilities), and those for which one or more of the `traceroute` hops was completely missing, as these measurements are inherently ambiguous. This left us with 31,709 measurements.

We next made a preliminary assessment of the patterns of route changes by seeing which occurred most frequently. We found the pattern of changes dominated by a number of single-hop differences, at which consecutive measurements showed exactly the same path except for an alternation at a single router. Furthermore, the names of these routers often suggested that the pair were administratively interchangeable. It seems likely that frequent route changes differing at just a single hop are due to shifting traffic between two tightly coupled machines. For the stability concerns given in § 7.1, such a change will have little consequence, provided the two routers are co-located and capable of sharing state. We identified 5 such pairs of "tightly coupled" routers and merged each pair into a single router for purposes of assessing stability (see [Pa96] for details).

Finally, we reduced the routes to three different levels of *granularity*: considering each route as a sequence of Internet hostnames (*host* granularity), as a sequence of cities (*city* granularity; see [Pa96] for details on geography), and as a sequence of AS's (*AS* granularity). The use of city and AS granularities introduces a notion of "major change" as opposed to "any change." Overall, 57% of the route changes at host granularity were also changes at city granularity, and 36% were changes at AS granularity.

7.4 Routing Prevalence

In this section we look at routing stability from the standpoint of *prevalence*: how likely we are, overall, to observe a particular route (c.f. § 7.2). We associate with prevalence a parameter π_r, the steady-state probability that a virtual path at an arbitrary point in time uses a particular route r.

We can assess π_r from our data as follows. We hypothesize that routing changes follow a semi-Markov process, in which case the steady-state probability of observing a particular state is equal to the average amount of time spent in that state [Ro83]. Because of PASTA, our sampling gives us exactly this time average (§ 4.2). So

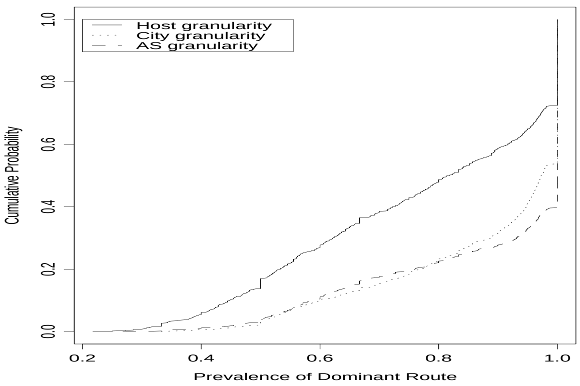

Figure 2: Fraction of observations finding the dominant route, for all virtual paths, at all granularities

if we make n observations of a virtual path and k_r of them find state r (i.e., route r), then we estimate $\hat{\pi}_r = k_r/n$.

For a particular virtual path p, let n_p be the total number of `traceroutes` measuring that virtual path, and k_p be the number of times we observed the *dominant* route, meaning the route that appeared most often. We focus our analysis on $\hat{\pi}_{\text{dom p}} = k_p/n_p$, the prevalence of the dominant route.

Figure 2 shows the cumulative distribution of the prevalence of the dominant routes over all 1,054 virtual paths measured in $\mathcal{D}_2$, for the three different granularities. There is clearly a wide range, particularly for host granularity. For example, for the virtual path between `pubnix` and `austr`, in 46 measurements we observed 9 distinct routes at host granularity, and the dominant route was observed only 10 times, leading to $\hat{\pi}_{\text{dom}} = 0.217$. On the other hand, at host granularity more than 25% of the virtual paths exhibited only a single route ($\hat{\pi}_{\text{dom}} = 1$). For city and AS granularities, the spread in $\hat{\pi}_{\text{dom}}$ is more narrow, as would be expected.

A key figure to keep in mind from this plot, however, is that while there is a wide range in the distribution of $\hat{\pi}_{\text{dom}}$ over different virtual paths, its *median* value at host granularity is 82%; 97% at city granularity, and 100% at AS granularity. Thus we can conclude: *In general, Internet paths are strongly dominated by a single route.*

Previous traffic studies, however, have shown that many characteristics of network traffic exhibit considerable site-to-site variation [Pa94], so it behooves us to assess the differences in $\hat{\pi}_{\text{dom}}$ between the sites in our study. To do so, for each site s (and for each granularity) we computed:

$$\hat{\pi}_{\text{src s}} = \sum_{\text{src virt-p } s_i} \frac{k_{s_i}}{n_{s_i}}.$$

Here "src virt-p" refers to all virtual paths that have their source at s. The aggregate estimate $\hat{\pi}_{\text{src s}}$ then indicates the overall prevalence of dominant routes from s to different destinations. We expect *variations* in this estimate for different sites to reflect differing routing prevalence due to route changes *near* the source. Route changes further downstream from the source occur either deep inside the network (and so will affect many different sites), or near the destination (and thus will not affect any particular *source* site unduly).

Similarly, we can construct $\hat{\pi}_{\text{dst s}}$ for all of the virtual paths with destination s. Studying $\hat{\pi}_{\text{src s}}$ and $\hat{\pi}_{\text{dst s}}$ for different sites and at different granularities reveals considerable site-to-site variation. For example, at host granularity, the prevalence of the dominant routes originating at the `ucl` source is under 50% (we will see why in § 7.5.1), and for `bnl`, `sintef1`, `sintef2`, and `pubnix` is around 60%, while for `ncar`, `ucol`, and `unij` it is just under 90%. Even at AS granularity, the `ucl` source has an average prevalence of 60%, with `ukc` about 70%, and the remainder from 85% to 99%. At city granularity the main outlier is `bnl`, with a prevalence of 75% (c.f. § 7.5.2), because the `ucl` and `ukc` instabilities, while spanning autonomous systems, do not span different cities.

We find similar spreads for $\hat{\pi}_{\text{dst s}}$. Some sites with low prevalence for $\hat{\pi}_{\text{src s}}$ have high prevalence for $\hat{\pi}_{\text{dst s}}$, and vice versa, due to *asymmetric* routing (§ 8).

We can thus summarize routing prevalence as follows: *In general, 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.*

7.5 Routing Persistence

We now turn to the more difficult task of assessing the *persistence* of routes: How long they are likely to endure before changing. As illustrated in § 7.2, routing persistence can be difficult to evaluate because a series of measurements at particular points in time do not necessarily indicate a lack of change *and then change back* in between the measurement points. Thus, to accurately assess persistence requires first determining if routing alternates on short time scales. If not, then we can trust shortly spaced measurements observing the same route as indicating that the route did indeed persist during the interval between the measurements. The shortly spaced measurements can then be used to assess whether routing alternates on medium time scales, etc. In this fashion, we aim to "bootstrap" ourselves into a position to be able to make sound characterizations of routing persistence across a number of time scales.

7.5.1 Rapid route alternation

We have already identified two types of rapidly alternating routes, those due to "flutter" and those due to "tightly coupled" routers. We have separately characterized fluttering (§ 6.4) and consequently have not included paths experiencing flutter in this analysis. As mentioned in § 7.3, we merged tightly coupled routers into a single entity, so their presence also does not further affect our analysis.

We next note that in $\mathcal{D}_2$ we observed 155 instances of a route change during a `traceroute`. The combined amount of time observed by the 35,109 $\mathcal{D}_2$ `traceroutes` was 881,578 seconds. (That is, the mean duration of a $\mathcal{D}_2$ `traceroute` was 25.1 seconds.) Since when observing the network for 881,578 seconds we saw 155 route changes, we can estimate that on average we will see a route change every 5,687 seconds ($\approx$ 1.5 hours). This reflects quite a high rate of route alternation, and bodes ill for relying on measurements made much more than a few hours apart (though see § 7.5.2); but not so high that we would expect to completely miss routing changes for sampling intervals significantly less than an hour.

We first looked at measurements made less than 60 seconds apart. There were only 54 of these, but all of them were of the form "R_1, R_1"—i.e., both measurements observed the same route. Thus there are no additional widespread, high-frequency routing oscillations.

We then looked at measurements made less than 10 minutes apart. There were 1,302 of these, and 40 *triple* observations (three

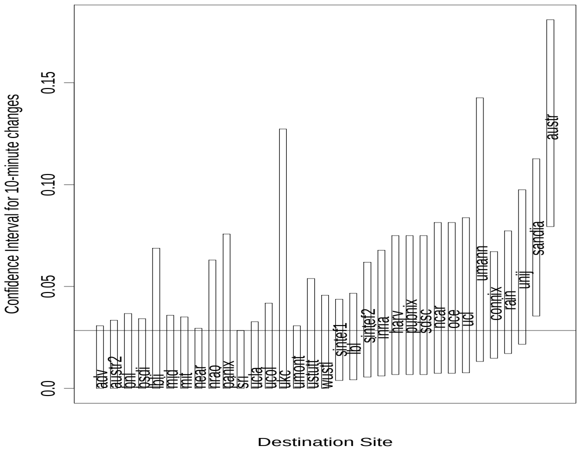

Figure 3: Site-to-site variation in $P^{10}_{\text{dst s}}$

observations all within a ten minute interval). The triple observations allow us to double check for the presence of high-frequency oscillations: if we observe the pattern R_1, R_2, R_1 or R_1, R_2, R_3, then we are likely to miss some route changes when using only two measurements 10 minutes apart. If we only observe R_1, R_1, R_1; R_1, R_2, R_2; or R_1, R_1, R_2, then measurements made 10 minutes apart are not missing short-lived routes. Of the 40 triple observations, all were of the latter forms.

The 1,302 ten-minute observations included 25 instances of a route change (R_1, R_2). This suggests that the likelihood of observing a route change over a ten minute interval is not negligible, and requires further investigation before we can look at more widely spaced measurements.

A natural question to ask concerning 10-minute changes is whether just a few sites are responsible for most of them. For each site s, let $N^{10}_{\text{src s}}$ be the number of 10-minute pairs of measurements originating at s, and $X^{10}_{\text{src s}}$ be the number of times those observed a change. Similarly, define $N^{10}_{\text{dst s}}$ and $X^{10}_{\text{dst s}}$ for those pairs of measurements with destination s. Here we are aggregating, for each site, all of the measurements made using that site as a source (destination), in an attempt to see whether route oscillations are significantly more prevalent near a handful of the sites.

For each site s, we can then define: $P^{10}_{\text{src s}} = X^{10}_{\text{src s}}/N^{10}_{\text{src s}}$, and similarly for $P^{10}_{\text{dst s}}$. $P^{10}_{\text{src s}}$ gives the estimated probability that a pair of ten-minute observations of virtual paths with source s will show a routing change. We can then use the methodology outlined in § 4.4 to associated confidence intervals with $P^{10}_{\text{src s}}$ and $P^{10}_{\text{dst s}}$, to see which sites, if any, exhibit significantly different probabilities of ten-minute changes.

Figure 3 shows the resulting confidence intervals for $P^{10}_{\text{dst s}}$. Sites are sorted according to the lower end of their confidence interval. Each interval is shown using a vertical bar, with the name of the site left-justified to start at $P^{10}_{\text{dst s}}$.

The horizontal line in the plot runs along the level corresponding to the smallest upper bound on $P^{10}_{\text{dst s}}$ (`sri`). *All sites with intervals intersecting the line are pairwise consistent with one another.* Those sites above the line (`sandia`, `austr`) are not consistent with the bulk of the other sites.

An important point here, however, is that the statistical comparison is valid for consistency between *pairs*. When plotting a whole set of confidence intervals, we must allow for a *multiplicity effect*: there is more opportunity for a few intervals to be inconsistent with the others, just due to chance. Thus, inconsistencies in the plot are *not* neccessarily significant. The plot *does*, however, point up outliers that merit further investigation. From this plot we conclude that `sandia` and (particularly) `austr` are outliers, much more likely (as destinations) subject to rapid routing oscillations. Before removing them as outliers, however, we must be careful to first look at their routing oscillations to see what patterns they exhibit.

For the destination `austr`, all of the changes (which involve a number of source sites) take place at the point-of-entry into Australia. The changes are either the first Australian hop of `vic.gw.au`, in Melbourne, or `act.gw.au`, in Canberra, or `serial4-6.pad-core2.sydney.telstra.net` in Sydney followed by an additional hop to `nsw.gw.au` (also in Sydney). These are the only points of change: before and after, the routes are unchanged. Thus, the destination `austr` exhibits rapid (time scale of tens of minutes) changes in its incoming routing. As such, the routing *to* `austr` is not at all persistent.

For `sandia`, however, the story is different. Its changes occurred only along the virtual path originating at `sri`, and reflected a change localized to MCINET in San Francsico. Had this change been more prevalent, we might have decided that the two pairs of routers in question were "tightly coupled" (§ 7.3), but they were responsible for changes only between `sri` and `sandia`. Thus, we can deal with this outlier by eliminating the virtual path `sri` ⇒ `sandia`, but keeping the other virtual paths with destination `sandia`.

In addition to the destination `austr`, a similar analysis of $P^{10}_{\text{src s}}$ points up `ucl`, `ukc`, `mid`, and `umann` as outliers. Both `ucl` and `ukc` had frequent oscillations in the routers visited between London and Washington, D.C., alternating between the two hops of:

```
icm-lon-1.icp.net, icm-dc-1-s3/2-1984k.icp.net
```

and the four hops of:

```
eu-gw.ja.net, gw.linx.ja.net,
us-gw.thouse.ja.net, icm-dc-1-s2/4-1984k.icp.net
```

Note that these different hops also correspond to different AS's, as the latter includes AS 786 (JANET) and the former does not. For `mid` and `umann`, however, the changes did not have a clear pattern, and their prevalence could be due simply to chance.

On the basis of this analysis, we conclude that the sources `ucl` and `ukc`, and the destination `austr`, suffer from significant, high-frequency oscillation, and excluded them from further analysis. After removing any measurements originating from the first two or destined to `austr`, we then looked at the range of values for $P^{10}_{\text{src s}}$ and $P^{10}_{\text{dst s}}$. Both of these had a median of 0 observed changes, and a maximum corresponding to about 1 change per hour. On this basis, we believe we are on firm ground treating pairs of measurements between these sites, made less than an hour apart, both observing the same route, as consistent with that route having persisted unchanged between the measurements.

7.5.2 Medium-scale route alternation

Given the findings that, except for a few sites, route changes do not occur on time scales less than an hour, we now turn to analyzing those measurements made an hour or less apart to determine what they tell us about medium-scale routing persistence. We proceed much as in § 7.5.1. Let $P^{\text{hr}}_{\text{src s}}$ and $P^{\text{hr}}_{\text{dst s}}$ be the analogs of $P^{10}_{\text{src s}}$ and $P^{10}_{\text{dst s}}$, but now for measurements made an hour or less apart. After eliminating the rapidly oscillating virtual paths previously identified, we have 7,287 pairs of measurements to assess.

The data also included 1,517 triple observations spanning an hour or less. Of these, only 10 observed the pattern R_1, R_2, R_1 or R_1, R_2, R_3, indicating that, in general, two observations spaced an hour apart are not likely to miss a routing change.

Plots similar to Figure 3 immediately pick out virtual paths originating from `bnl` as exhibiting rapid changes. These changes are almost all from oscillation between `llnl-satm.es.net` and `pppl-satm.es.net`. (The first in California, the second in New Jersey). ESNET oscillations also occurred on one-hour time scales in traffic between `lbl` (and `lbli`) and the Cambridge sites, `near`, `harv`, and `mit`.

The other prevalent oscillation we found was between the source `umann` and the destinations `ucl` and `ukc`. Here the alternation was between a British Telecom router in Switzerland and another in the Netherlands.

Eliminating these oscillating virtual paths leaves us with 6,919 measurement pairs. These virtual paths are not statistically identical (i.e., we find among them paths that have significantly different route change rates), but all have low rates of routing changes. For these virtual paths, the median $P^{\mathrm{hr}}_{\mathrm{src\ s}}$ and $P^{\mathrm{hr}}_{\mathrm{dst\ s}}$ correspond to one routing change per 1.5 days, and the maximum to one change per 12 hours.

7.5.3 Large-scale route alternation

Given that, after removing the oscillating paths discussed above, we expect at most on the order of one route change per 12 hours, we now can analyze measurements less than 6 hours apart of the remaining virtual paths to assess longer-term route changes. There were 15,171 such pairs of measurements. As 6 hours is significantly larger than the mean 2 hour sampling interval, not surprisingly we find many triple measurements spanning less than 6 hours. But of the 10,660 triple measurements, only 75 included a route change of the form R_1, R_2, R_1 or R_1, R_2, R_3, indicating that, for the virtual paths to which we have now narrowed our focus, we are still not missing many routing changes using measurements spaced up to 6 hours apart.

Employing the same analysis, we first identify `sintef1` and `sintef2` as outliers, both as source and as destination sites. The majority of their route changes turn out to be oscillations between two sets of routers, each alternating between visiting or not visiting Oslo. Two other outliers at this level are traffic to or from `sdsc`, which alternates between two different pairs of CERFNET routers in San Diego, and traffic originating from `mid`, which alternates between two MIDNET routers in St. Louis.

Eliminating these paths leaves 11,174 measurements of the 712 remaining virtual paths. The paths between the sites in these remaining measurements are quite stable, with a maximum transition rate for any site of about one change every two days, and a median rate of one per four days.

7.5.4 Duration of long-lived routes

We term the remaining measurements as corresponding to "long-lived" routes. For these, we might hazard to estimate the durations of the different routes as follows. We suppose that we are not completely missing any routing transitions, an assumption based on the overall low rate of routing changes. Then for a sequence of measurements all observing the same route, we assume that the route's

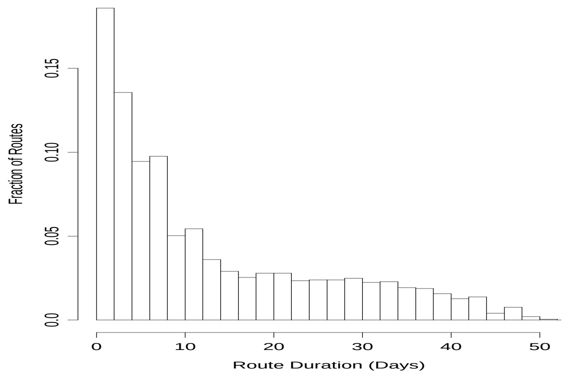

Figure 4: Estimated distribution of long-lived route durations

Time scale	%	Notes
seconds	N/A	"Flutter" for purposes of load balancing. Treated separately, as a pathology, and not included in the analysis of persistence.
minutes	N/A	"Tightly-coupled routers." We identified five instances, which we merged into single routers for the remainder of the analysis.
10's of minutes	9%	Frequent route changes inside the network. In some cases involved routing through different cities or AS's.
hours	4%	Usually intra-network changes.
6+ hours	19%	Also intra-network changes.
days	68%	Bimodal. 50% of routes persist for under 7 days. The remaining 50% account for 90% of the total route lifetimes.

Table 4: Summary of persistence at different time scales

duration was at least the span of the measurements. Furthermore, if at time t_1 we observe route R_1 and then the next measurement at time t_2 observes route R_2, we make a "best guess" that route R_1 terminated and route R_2 began half way between these measurements, i.e., at time $\frac{t_1+t_2}{2}$. (See [Pa96] for additional details.)

Figure 4 shows the distribution of the estimated durations of the "long-lived" routes. Even keeping in mind that our estimates are rough, it is clear that the distribution of long-lived route durations has two distinct regions, with many of the routes persisting for 1-7 days, and another group persisting for several weeks. About half the routes persisted for under a week, but the half of the routes lasting more than a week accounted for 90% of total persistence. This means that if we observe a virtual path at an arbitrary point in time, *and we are not observing one of the numerous, more rapidly oscillating paths outlined in the previous sections*, then we have about a 90% chance of observing a route with a duration of at least a week.

7.5.5 Summary of routing persistence

We summarize routing persistence as follows. First, *routing changes occur over a wide range of time scales, ranging from seconds to days.* Table 4 lists different time scales over which routes change. The second column gives the percentage of all of our measured virtual paths (source/destination pairs) that were affected by changes at the given time scale. (The first two rows show "N/A" in this field because the changes were due to a very small, and hence not representative, set of routers.) The final column gives associated notes.

One important point apparent from the table is that routing changes on shorter time scales (fewer than days) happen *inside the network* and not at the stub networks. Thus, *those changes observed in our measurements are likely to be similar to those observed by most Internet sites*.

On the other hand, while the changes occurred inside the network, only those involving `ucl` and `ukc` (§ 7.5.1) involved different sequences of autonomous systems. While this bodes well for the scalability of BGP, we do not claim this finding as having major significance: one could make a much more thorough assessment of the degree of inter-AS route flapping by analyzing the data discussed in [Do95, Me95b].

Finally, two thirds of the Internet paths we studied had quite stable routes, persisting for days or weeks. This finding is in accord with that of [Ch93], which found that most networks are nearly quiescent (in terms of routing changes) while a few exhibit frequent connectivity transitions.

8 Routing symmetry

We now analyze the measurements to assess the degree to which routes are *symmetric* or *asymmetric*. We confine ourselves to studying "major" asymmetries, in which the sequence of cities or AS's visited by the routes for the two directions of a virtual path differ. We first discuss the impact of routing asymmetry on different network protocols and measurements. We then assess our data for these asymmetries and find that, overall, 50% of the time an Internet path includes a major asymmetry in terms of the cities visited in the different directions, and 30% of the time it includes a major asymmetry in terms of AS's visited. We finish with a discussion of the magnitude of the asymmetries, most of which differ at just one "hop," but some at many hops.

8.1 Importance of routing symmetry

Routing symmetry affects a number of aspects of network behavior. When attempting to assess the one-way propagation time between two Internet hosts, the common practice is to assume it is well approximated as half of the round-trip time (RTT) between the hosts [CPB93]. The Network Time Protocol (NTP) needs to make such an assumption when synchronizing clocks between widely separated hosts [Mi92].[2]

Claffy and colleagues studied variations in one-way latencies between the United States, Europe, and Japan [CPB93]. They discuss the difficulties of measuring *absolute* differences in propagation times in the absence of separately-synchronized clocks, but for their study they focussed on *variations*, which does not require synchronization of the clocks. They found that the two opposing directions of a path do indeed exhibit considerably different latencies, in part due to different congestion levels, and in part due to routing changes.

Routing asymmetry also potentially complicates mechanisms by which endpoints infer network conditions from the pattern of packet arrivals they observe, and the utility of routers establishing *anticipatory flow state* when they observe a new flow from A to B that is likely to generate a return flow from B to A [CBP95]. See [Pa96] for detailed discussion of these.

Finally, routing asymmetry complicates network troubleshooting, because it increases the likelihood that a network problem apparent in one direction along a virtual path cannot be detected in the other direction.

8.2 Analysis of routing symmetry

In $\mathcal{D}_1$ we did not make simultaneous measurements of the virtual paths $A \Rightarrow B$ and $B \Rightarrow A$, which introduces ambiguity into an analysis of routing symmetry: if a measurement of $A \Rightarrow B$ is asymmetric to a later measurement of $B \Rightarrow A$, is that because the route is the same but asymmetric, or because the route changed?

In $\mathcal{D}_2$, however, the bulk of the measurements were *paired* (§ 4.1), allowing us to unambiguously determine whether the route between A and B is symmetric. The $\mathcal{D}_2$ measurements contain 11,339 successful pairs of measurements. Of these, we find that *49% of the measurements observed an asymmetric path that visited at least one different city*.

There is a large range, however, in the prevalence of asymmetric routes among virtual paths to and from the different sites. For example, 86% of the paths involving `umann` were asymmetric, because nearly all outbound traffic from `umann` traveled via Heidelberg, but none of the inbound traffic did. At the other end of the spectrum, only 25% of the paths involving `umont` were asymmetric (but this is still a significant amount).

If we consider autonomous systems rather than cities, then we still find asymmetry quite common: about 30% of the paired measurements observed different autonomous systems in the virtual path's two directions. The most common asymmetry was the addition of a single AS in one direction. This can reflect a major change, however, such as the presence or absence of SprintLink routers (the most common AS change).

Again, we find wide variation in the prevalence of asymmetry among the different sites. Fully 84% of the paths involving `ucl` were asymmetric, mostly due to some paths including JANET routers in London and others not (§ 7.5.1), while only 7.5% of `adv`'s paths were asymmetric at AS granularity.

8.3 Increasing prevalence of asymmetry

We analyzed $\mathcal{D}_1$ for routing asymmetry, attempting to adjust for the non-simultaneity of its measurements by only using measurements spaced less than a day apart. The mismatch is likely to overestimate routing asymmetry, since if the route changes between measurements that may be incorrectly regarded as an asymmetry, per our discussion at the beginning of § 8.2, though it can also introduce false symmetries.

In the $\mathcal{D}_1$ measurements, we found 30% of the virtual paths contained city-level asymmetries. The large discrepancy between this figure and the 49% figure for the $\mathcal{D}_2$ measurements suggests that over the course of a year routing became significantly more asymmetric.

8.4 Size of asymmetries

We finish with a look at the size of the asymmetries. We find that the majority of asymmetries are confined to a single "hop" (just one

[2] However, NTP features robust algorithms that will only lead to inconsistencies if the paths between two NTP communities are *predominantly* asymmetric, with similar differences in one-way times.

city or AS different). For city asymmetries, though, about one third differed at two or more "hops." This corresponds to almost 20% of all the paired measurements in our study, and can indicate a very large asymmetry. For example, a magnitude 2 asymmetry between `ucl` and `umann` differs at the central city hops of Amsterdam and Heidelberg in one direction, and Princeton and College Park in the other!

9 Summary

We have reported on an analysis of 40,000 end-to-end Internet route measurements, conducted between a diverse collection of Internet sites. The study characterizes pathological routing conditions, routing stability, and routing symmetry. For pathologies, we found a number of examples of routing loops, some persisting for hours; one instance of erroneous routing; a number of instances of "infrastructure failures," meaning that routing failed deep inside the network; and numerous outages lasting 30 seconds or more.

Our statistical methodology allows us to assign confidence intervals to the probabilities of observing different pathologies, and to compare these intervals for significant differences. We find that *the likelihood of encountering a major routing pathology more than doubled between the end of 1994 and the end of 1995, rising from 1.5% to 3.4%.*

For routing stability, we defined two types of stability, "prevalence," meaning the overall likelihood that a particular route is encountered, and "persistence," the likelihood that a route remains unchanged over a long period of time. We find that *Internet paths are heavily dominated by a single prevalent route, but that the time periods over which routes persist show wide variation*, ranging from seconds up to days. About 2/3's of the Internet paths had routes persisting for either days or weeks.

For routing symmetry, we looked at the likelihood that a virtual path through the Internet visits at least one different city in the two directions. At the end of 1995, this was the case half the time, nearly double the likelihood at the end of 1994, and at least one different autonomous system was visited 30% of the time.

The presence of pathologies, short-lived routes, and major asymmetries highlights the difficulties of providing a consistent topological view in an environment as large and diverse as the Internet. Furthermore, the findings that the prevalence of pathologies and asymmetries greatly increased during 1995 show in no uncertain terms that *Internet routing has become less predictable in major ways*.

A constant theme running through our study is that of widespread variation. We repeatedly find that different sites or pairs of sites encounter very different routing characteristics. This finding matches that of [Pa94], which emphasizes that the variations in Internet traffic characteristics between sites are significant to the point that there is no "typical" Internet site. Similarly, there is no "typical" Internet path. But we believe the scope of our measurements gives us a solid understanding of the breadth of behavior we might expect to encounter—and how, from an end-point's view, routing in the Internet actually works.

10 Acknowledgements

This work would not have been possible without the efforts of the many volunteers who installed the Network Probe Daemon at their sites. In the process they endured debugging headaches, `inetd` crashes, software updates, and a seemingly endless stream of queries from me regarding their site's behavior. I am indebted to:

> 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, 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`).

I am likewise indebted to Keith Bostic, Evi Nemeth, Rich Stevens, George Varghese, Andres Albanese, Wieland Holfelder, and Bernd Lamparter for their invaluable help in recruiting NPD sites. Thanks too to Peter Danzig, Jeff Mogul, and Mike Schwartz for feedback on the NPD design.

This work greatly benefited from the efforts and insights of Domenico Ferrari, Sally Floyd, John Hawkinson, Van Jacobson, Kurt Lidl, Steve McCanne, Greg Minshall, Craig Partridge, and John Rice, all of whom gave detailed comments on [Pa96]; from discussions with Guy Almes, Robert Elz, Teus Hagen, John Hawkinson, Kate Lance, Paul Love, Jamshid Mahdavi, Matt Mathis, Dave Mills, and Curtis Villamizar; and from the comments of the SIGCOMM reviewers. I would like to particularly thank Van Jacobson and John Rice for their advice regarding the thorny problem of assessing routing "persistence."

Often to understand the behavior of particular routers or to determine their location, I asked personnel from the organization responsible for the routers. I was delighted at how willing they were to help, and in this regard would like to acknowledge:

> Vadim Antonov, Tony Bates, Michael Behringer, Per Gregers Bilse, Bjorn Carlsson, Peggy Cheng, Guy Davies, Sean Doran, Bjorn Eriksen, Amit Gupta, Tony Hain, John Hawkinson again!, Susan Harris, Ittai Hershman, Kevin Hoadley, Scott Huddle, James Jokl, Kristi Keith, Harald Koch, Craig Labovitz, Tony Li, Martijn Lindgreen, Ted Lindgreen, Dan Long, Bill Manning, Milo Medin, Keith Mitchell, Roderik Muit, Chris Myers, Torben Nielsen, Richard Nuttall, Mark Oros, Michael Ramsey, Juergen Rauschenbach, Douglas Ray, Brian Renaud, Jyrki Soini, Nigel Titley, Paul Vixie, and Rusty Zickefoose.

A preliminary analysis of $\mathcal{D}_1$ was done by Mark Stemm and Ketan Patel.

References

[Ba95] F. Baker, Ed., "Requirements for IP Version 4 Routers," RFC 1812, DDN Network Information Center, June 1995.

[BDG95] C. Baransel, W. Dobosiewicz, and P. Gburzynski, "Routing in Multihop Packet Switching Networks: Gb/s Challenge," *IEEE Network*, 9(3), pp. 38-61, May/June 1995.

[BM92] I. Bilinskis and A. Mikelsons, *Randomized Signal Processing*, Prentice Hall International, 1992.

[Bo93] J-C. Bolot, "End-to-End Packet Delay and Loss Behavior in the Internet," *Proceedings of SIGCOMM '93*, pp. 289-298, September 1993.

[BCS94] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview," RFC 1633, DDN Network Information Center, June 1994.

[BE90] L. Breslau and D. Estrin, "Design of Inter-Administrative Domain Routing Protocols," *Proceedings of SIGCOMM '90*, pp. 231-241, September 1990.

[Ch93] B. Chinoy, "Dynamics of Internet Routing Information," *Proceedings of SIGCOMM '93*, pp. 45-52, September 1993.

[CBP95] K. Claffy, H-W. Braun and G. Polyzos, "A Parameterizable Methodology for Internet Traffic Flow Profiling," *IEEE JSAC*, 13(8), pp. 1481-1494 October 1995.

[CPB93] K. Claffy, G. Polyzos and H-W. Braun, "Measurement Considerations for Assessing Unidirectional Latencies," *Internetworking: Research and Experience*, 4 (3), pp. 121-132, September 1993.

[DC90] S. Deering and D. Cheriton, "Multicast Routing in Datagram Internetworks and Extended LANs," *ACM Transactions on Computer Systems*, 8(2), pp. 85-110, May 1990.

[DB95] L. Delgrossi and L. Berger, Ed., "Internet Stream Protocol Version 2 (ST2), Protocol Specification—Version ST2+," RFC 1819, DDN Network Information Center, August 1995.

[Do95] Sean Doran, "Route Flapping," with notes by Stan Barber, *http://www.merit.edu/routing.arbiter/NANOG/2.95.NANOG.notes/route-flapping.html.*

[ERH92] D. Estrin, Y. Rekhter and S. Hotz, "Scalable Inter-Domain Routing Architecture," *Proceedings of SIGCOMM '92*, pp. 40-52, August 1992.

[FBZ94] D. Ferrari, A. Banerjea and H. Zhang, "Network support for multimedia: A discussion of the Tenet approach," *Computer Networks and ISDN Systems*, 26(10), pp. 1267-1280, July 1994.

[FJ94] S. Floyd and V. Jacobson, "The Synchronization of Periodic Routing Messages," *IEEE/ACM Transactions on Networking*, 2(2), pp. 122-136, April 1994.

[HK89] S. Hares and D. Katz, "Administrative Domains and Routing Domains: A Model for Routing in the Internet," RFC 1136, Network Information Center, SRI International, Menlo Park, CA, December, 1989.

[Hu95] C. Huitema, *Routing in the Internet*, Prentice Hall PTR, 1995.

[Ja89] V. Jacobson, *traceroute*, *ftp://ftp.ee.lbl.gov/traceroute.tar.Z*, 1989.

[KZ89] A. Khanna and J. Zinky, "The Revised ARPANET Routing Metric," *Proceedings of SIGCOMM '89*, pp. 45-56, 1989.

[La96] C. Labovitz, private communication, May 1996.

[Li89] M. Little, "Goals and Functional Requirements for Inter-Autonomous System Routing," RFC 1126, Network Information Center, SRI International, Menlo Park, CA, October, 1989.

[Lo95] M. Lottor, *ftp://nic.merit.edu/nsfnet/statistics*; October, 1995.

[MFR78] J. McQuillan, G. Falk and I. Richer, "A Review of the Development and Performance of the ARPANET Routing Algorithm," *IEEE Transactions on Communications*, 26(12), pp. 1802-1811, December 1978.

[MRR80] J. McQuillan, I. Richer and E. Rosen, "The New Routing Algorithm for the ARPANET," *IEEE Transactions on Communications*, 28(5), pp. 711-719, May 1980.

[Me95a] Merit Network, Inc., *ftp://nic.merit.edu/nsfnet/statistics/history.nets*; May, 1995.

[Me95b] Merit Network, Inc., *http://nic.merit.edu/routing.arbiter/RA/statistics/flap.html*.

[Mi92] D. Mills, "Network Time Protocol (Version 3): Specification, Implementation and Analysis," RFC 1305, Network Information Center, SRI International, Menlo Park, CA, March 1992.

[Mu94] A. Mukherjee, "On the Dynamics and Significance of Low Frequency Components of Internet Load," *Internetworking: Research and Experience*, Vol. 5, pp. 163-205, December 1994.

[Pa94] V. Paxson, "Empirically-Derived Analytic Models of Wide-Area TCP Connections," *IEEE/ACM Transactions on Networking*, 2(4), pp. 316-336, August 1994.

[Pa96] V. Paxson, Part I of *An Analysis of End-to-End Internet Dynamics,* Ph.D. dissertation in preparation, University of California, Berkeley, 1996.

[PV88] R. Perlman and G. Varghese, "Pitfalls in the Design of Distributed Routing Algorithms," *Proceedings of SIGCOMM '88*, pp. 43-54, August 1988.

[Pe91] R. Perlman, "A comparison between two routing protocols: OSPF and IS-IS," *IEEE Network*, 5(5), pp. 18-24, September 1991.

[Pe92] R. Perlman, *Interconnections: Bridges and Routers,* Addison-Wesley, 1992.

[RC92] Y. Rekhter and B. Chinoy, "Injecting Inter-autonomous System Routes into Intra-autonomous System Routing: a Performance Analysis," *Internetworking: Research and Experience*, Vol. 3, pp. 189-202, 1992.

[Re95] Y. Rekhter, "Inter-Domain Routing: EGP, BGP, and IDRP," in [St95].

[RL95] Y. Rekhter and T. Li, "A Border Gateway Protocol 4 (BGP-4)," RFC 1771, DDN Network Information Center, March 1995.

[RG95] Y. Rekhter and P. Gross, "Application of the Border Gateway Protocol in the Internet," RFC 1772, DDN Network Information Center, March 1995.

[Ro82] E. Rosen, "Exterior Gateway Protocol (EGP)," RFC 896, Network Information Center, SRI International, Menlo Park, CA, October 1982.

[Ro83] S. Ross, *Stochastic Processes*, John Wiley & Sons, 1983.

[SAGJ93] D. Sanghi, A.K. Agrawal, Ó. Gudmundsson, and B.N. Jain, "Experimental Assessment of End-to-end Behavior on Internet," *Proceedings of INFOCOM '93*, San Francisco, March, 1993.

[SS80] M. Schwartz and T. Stern, "Routing Techniques Used in Computer Communication Networks," *IEEE Transactions on Communications*, 28(4), pp. 539-552, April 1980.

[SFANC93] D. Sidhu, T. Fu, S. Abdallah, R. Nair, and R. Coltun, "Open Shortest Path First (OSPF) Routing Protocol Simulation," *Proceedings of SIGCOMM '93*, pp. 53-62, September 1993.

[St95] M. Steenstrup, editor, *Routing in Communications Networks*, Prentice-Hall, 1995.

[St94] W.R. Stevens, *TCP/IP Illustrated, Volume 1: The Protocols*, Addison-Wesley, 1994.

[Tr95a] P. Traina, "Experience with the BGP-4 Protocol," RFC 1773, DDN Network Information Center, March 1995.

[Tr95b] P. Traina, editor, "BGP-4 Protocol Analysis," RFC 1774, DDN Network Information Center, March 1995.

[Wo82] R. Wolff, "Poisson Arrivals See Time Averages," *Operations Research*, 30(2), pp. 223-231, 1982.

[ZG-LA92] W. Zaumen and J.J. Garcia-Luna Aceves, "Dynamics of Link-state and Loop-free Distance-vector Routing Algorithms," *Internetworking: Research and Experience*, Vol. 3, pp. 161-188, 1992.

[ZDESZ93] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: A New Resource ReSerVation Protocol," *IEEE Network*, 7(5), pp. 8-18, September 1993.