The spectrum of Internet performance

Luca Carbone*, Fabrizio Coccetti†, Paolo Dini*, Roberto Percacci§, Alessandro Vespignani¶

* Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
† Spin s.r.l., Via del Follatoio 12, 34148 Trieste and Stanford Linear Accelerator Center, 94025 Menlo Park, California
§ SISSA, Via Beirut 4, 34014 Trieste and Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Italy
¶ Laboratoire de Physique Theorique, Batiment 210, Universite de Paris-Sud, 91405 ORSAY Cedex

***Abstract*— We advocate the use of certain dimensionless metrics, depending on the ratio of RTT over distance, to measure the quality of Internet paths. We have analyzed several independent samples of Internet paths, finding that the statistical distributions of these metrics are quite simple and show power-law tails. Understanding the exponents of these tails is of special interest to model-building. We exhibit the results of numerical simulations on very simple models that reproduce to some extent the general features of these distributions. While not yet sufficiently realistic, these models suggest some of the mechanisms that may be shaping the large scale behaviour of the Internet.**

I. Introduction

The most widely used metric of Internet performance is Round Trip Time (RTT). For certain purposes knowledge of the absolute value of RTT is necessary, but in judging the quality of an Internet path between two sites it makes more sense to take into account also the distance between the sites. For example, an RTT of 60ms would be good for a transatlantic link, but very bad for a London-Paris link. Thus, as a measure of the quality of an Internet path, RTT relative to distance is more significant than RTT itself.

Let us define $d = 2D/v$, where D is the geographic (great circle) distance between to sites, v is the speed of light in optical fiber (approximately 2/3 of the speed of light in vacuum) and the factor 2 is there to take into account the fact that a measure of RTT involves travelling twice the distance between the two sites. Numerically, the conversion factor $2/v$ is assumed equal to 1ms/100km in this paper. d is essentially a measure of distance in light-milliseconds, and it is an absolute physical lower bound on RTT. By a slight abuse of language we will refer to d as "the distance" in the rest of this paper.

The dimensionless quantity

$$\tau = \mathrm{RTT}/d \tag{1}$$

can be taken as a measure of the (instantaneous) quality of an Internet path. If the path consisted of a single, unloaded, straight link, τ would be equal to one. In the real world, τ will always be larger than one, and the larger it is the worse the performance of the path relative to the best theoretically possible performance. Since d does not change with time, τ is simply proportional to RTT. On a given link, it is a function of time. Repeated measurements of RTT (and hence of τ) will yield a whole distribution of values. These distributions have been analyzed for instance in [1]. Rather than look at the distribution of τ over time for a single path, we will consider only the monthly minimum and average values of τ for a path, and look at the distribution of these quantities for a large number of paths. These distributions describe what may be called the spectrum of Internet performance.

What do these distributions tell us about the network? There are several factors contributing to the RTT between two sites. Since an Internet path consists of links and nodes, one can roughly divide them into propagation times (the time it takes the signals to travel along the links) and processing times (the time it takes the nodes to read, route and resend the data). See [2] for a careful discussion. Modern routers, especially the ones used in backbones, have very low processing times and therefore, in the absence of congestion, and when the sites are not too close, the most important single contribution to RTT is the propagation time. Assuming that all links have the same propagation speed v, the contribution of propagation time to RTT is equal to $\ell = 2L/v$, where L is the total length of the physical path travelled by the signal. A network is likely to be congestion-free at least at some times in a daily or weekly cycle. At those times, RTT can be taken as an estimate of ℓ. Therefore, denoting $\mathrm{RTT}_{\mathrm{min}}$ the minimum RTT observed on a link over a sufficiently long period,

$$\tau_{\mathrm{min}} = \mathrm{RTT}_{\mathrm{min}}/d \tag{2}$$

is an estimate of the purely geometric ratio ℓ/d, i.e. a measure of how much the physical path travelled by the signal deviates from being straight. We may call this ratio the wiggliness of the path.

On the other hand, the ratio

$$\tau_{\mathrm{avg}} = \mathrm{RTT}_{\mathrm{avg}}/d \tag{3}$$

is sensitive to both propagation and processing delays. Insofar as processing delays can be neglected when the network is unloaded, processing delays arise from queueing at the egress interfaces of the routers and are therefore a symptom of congestion. Thus, τ_{avg} is a measure of a path's performance that takes into account both the wiggliness of, and the congestion on the path.

In an empty network, $\mathrm{RTT}_{\mathrm{avg}}/\mathrm{RTT}_{\mathrm{min}}$ would be equal to one, so the quantity

$$\rho = \tau_{\mathrm{avg}}/\tau_{\mathrm{min}} = \mathrm{RTT}_{\mathrm{avg}}/\mathrm{RTT}_{\mathrm{min}} \tag{4}$$

is a measure of the presence of queueing delays on the network, independent of the geometry of the data path. The

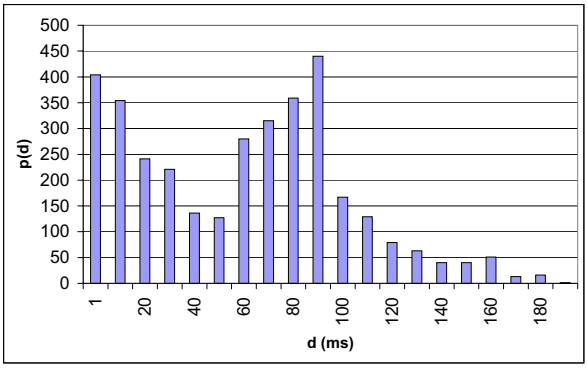

Fig. 1. Probability density distribution of distance in the PingER sample (2002).

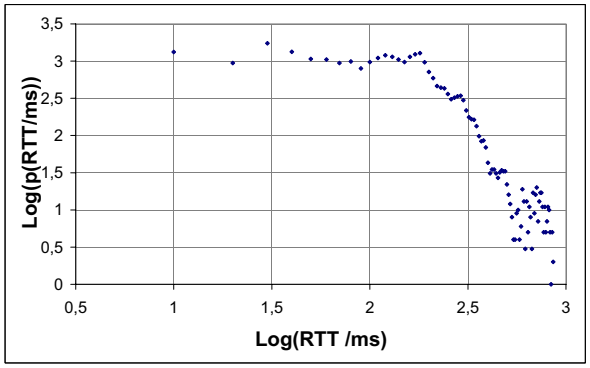

Fig. 2. Probability density distribution of RTT in the PingER sample (2002), in log-log scale.

value of ρ for a particular link gives us some information on how congested the link is, on average.

There is an obvious difference between these metrics, that is $\tau_{\min}$ is essentially constant over time (except for possible changes in routing) whereas τ_{avg} and ρ change continuously due to the changing traffic patterns. Therefore $\tau_{\min}$ is an almost constant, essentially deterministic variable, while τ_{avg} and ρ only give the time averages of quantities that fluctuate stochastically.

We claim that $\tau_{\min}$, τ_{avg} and ρ are more significant than RTT as measures of network quality. Being dimensionless is natural for a quality parameter. Furthermore, these quantities also have distributions that are much nicer than the ones of the dimensionful quantities that they are constructed from. For example, Fig. 1 shows the distribution of d in the PingER sample (to be discussed more extensively in Section III) on a linear scale. It is a complicated distribution reflecting the actual sizes of the continents and oceans. Due to the fact that North America and Europe are approximately 4000km across, there is a dearth of paths with distances of the order of 5000km, while most intercontinental paths are in the range of 10000km. On the other hand, Fig. 2 shows the distribution of $\text{RTT}_{\min}$ for the same sample (in a log-log scale). It is almost flat for low delays, has a sharp bend at 190ms and decays fast above that value. Fig. 3 shows the distribution of the ratio $\tau_{\min} = \text{RTT}_{\min}/d$ in log-log scale. It is seen to have a much nicer distribution than the other two quantities, with a clean peak at $\tau_{\min}$ =1.8 and a power law tail [3]. This distribution is unaffected by accidents of geography, since a given value of $\tau_{\min}$ can be due to hosts having any geographic distance. It therefore represents a more intrinsic property of the network.

In this connection, note that a high value of $\tau_{\min}$ (and therefore, presumably, of the ratio ℓ/d) is not necessarily due to a high value of $\text{RTT}_{\min}$: it could also be due to a relatively small value of d. For example, the path between two hosts that are located in the same building but are connected to two

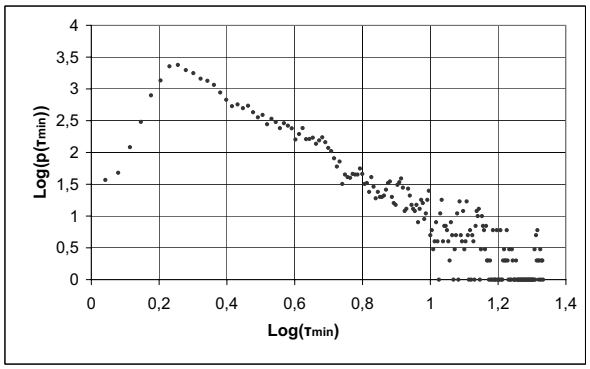

Fig. 3. Probability density distribution of $\tau_{\min}$ in the PingER sample (2002), in log-log scale.

different ISPs could easily have a value of $\tau_{\min}$ of the order of thousands. By putting these hosts right next to each other we could produce even larger values of $\tau_{\min}$, with a cutoff roughly equal to (size of the host)/(size of the network) which is of the order of millions. Thus, for practical purposes, $\tau_{\min}$ and τ_{avg} are essentially unbounded.

Very high values of these variables certainly arise in the real world, but they are not present in the samples we have examined. In fact, due to the difficulty of knowing the distance of the hosts when they are in the same city, and to the uncertainties of sub-millisecond measurements of RTT, in our analyses we have neglected all host pairs that are less that 1ms (100km) apart.

II. Related work

Before describing our results in detail, we comment on related work. Many people have made remarks that are more or less related to the subject of this paper, but to the best of our knowledge there have not been many systematic analyses

of the relation between RTT and distance.

Huffaker *et al.* [4] have given many interesting plots (in particular scatterplots of RTT vs distance) taken from some skitter monitors. Lee and Stepanek [5], in order to evaluate the impact of network performance on GRID infrastructure, have also given scatterplots involving RTT, distance and throughput for a sample of data collected within the Gloperf project. Their sample is discussed in Sect III. Bovy *et. al* [2] have considered a small sample of paths between RIPE TTM test-boxes and have given a detailed analysis of the various components contributing to the RTT. Subramanian *et al.* [6] have also discussed the relation between RTT and distance. They give distributions of the ratio ℓ/d for a very large sample of paths. They also discuss in detail several anecdotal cases that shed light on the origin of high values of this ratio. Unfortunately, they do not give a mathematical description of the tails of the distributions. It will be interesting to compare them with the ones we observe.

Many of these works, when estimating the length of a path, use traceroute techniques. Measuring the length of an IP path by summing the geographic distances between the nodes along the path underestimates the wiggliness ℓ/d, for two reasons. First, this technique implicitly assumes that the level-two links between nodes are straight, which is rarely true; second, if a node cannot be identified it is usually neglected and this (by the triangle inequalities) lowers the estimate for ℓ.

In [3] and in the present paper, we take $\mathrm{RTT}_{\min}$ as an estimate for ℓ. In doing so, we overestimate ℓ because the measured value of $\mathrm{RTT}_{\min}$ also includes the effect of processing delays, which we ignore. Thus, the traceroute method always gives a lower bound for ℓ/d, whereas our method always gives an upper bound. The statement in [6] that $\mathrm{RTT}_{\min}$ and the sum of the distances between nodes are strongly correlated is an encouraging indication that both methods yield results that are quite close to the truth.

In general, we would like to stress that our approach is to consider only large scale, average properties of the Internet. This precludes the possibility of considering too many details of each path. We believe that our method, being very simple, has the greatest chance of becoming a practical way of measuring network performance on a global scale.

III. Data Analyses

There are several active measurement infrastructures that regularly monitor the RTT between geographically known sites. We have considered four independent data sets, gathered by PingER [7], by the Gloperf experiment [5], by the RIPE TTM infrastructure [9] and by NLANR's AMP infrastructure [8]. For each data set we have worked out the distributions of the quantities $\tau_{\min}$, τ_{avg} and ρ, and for some also of Packet Loss (PL).

The main result of these analyses is that the tails of all these distributions can be approximated (with varying degrees of accuracy) by power laws:

$$p(t) = t^{-\alpha} \tag{5}$$

where the exponent alpha does not change too much from one sample to another.

Since the tails of the distributions become quite noisy towards the end, in order to smooth out fluctuations we will often plot the cumulative (integral) distributions $P(t)$. They are related to the distribution $p(t)$ by

$$P(t) = \int_t^\infty p(t')dt', \tag{6}$$

e.g. in the case of $\tau_{\min}$, $P(t)$ is the number of paths with $\tau_{\min}$ larger than t. If the probability density distribution $p(t)$ has the form (5), the graph of the cumulative distribution $P(t)$ in log-log scale is a straight line with slope $-\alpha+1$. ¿From these graphs one can usually better judge the quality of the power-law fit.

The distributions we show are not normalized. For visual clarity we have shifted some of the graphs vertically in some of the figures. This is equivalent to multiplying $P(t)$ by a constant and is therefore equivalent to changing the total number of events in the sample.

The following subsections describe the results of each data set. Readers who are only interested in the final results can go straight to the comparative summary of all the data, given in Table II below.

A. PingER data

PingER is a project of the IEPM group at SLAC [10] that has has been continuously collecting ping data since 1995. For an overview of the project see [7]. We have analyzed data taken from april 1999 to december 2002, covering almost three years. The sample we have analyzed consists of over 4000 Internet paths originating from 36 monitoring sites in 14 countries and targeting two hundred hosts in 77 countries. For every path, PingER makes roughly 1450 measurements in the course of a month, each consisting of a train of 21 100Byte ICMP packets. From these 1450 measurements the minimum and the average RTT, as well as the value of Packet Loss, are extracted. From the known geographical positions of the hosts we have computed the geographic distance and produced monthly tables of $\tau_{\min}$, τ_{avg} and PL, each containing roughly 2000 data points (not all paths are active every month). Some of the paths are also tested with 1000-Byte pings. In any given month there are approximately 1500 such data points. We display the results of the 100- and 1000-Byte packets in the same figures.

We begin by presenting the overall data of the year 2002. The cumulative distribution $P(\tau_{\min})$, in log-log scale is shown in Fig. 4. The corresponding probability density distribution (for 100Byte packets) was shown in Fig. 3. The plateau for small taumin corresponds to the region below the peak of the distribution in Fig. 3 at $\tau_{\min}$=1.8, and there is a clear cutoff at $\tau_{\min}$=14. In between lies a region spanning approximately one order of magnitude which is very well approximated by a straight line with slope -2.2. This means that the tail of the distribution $p(\tau_{\min})$ in Fig. 3 has an exponent $\alpha_{\min}$=3.2.

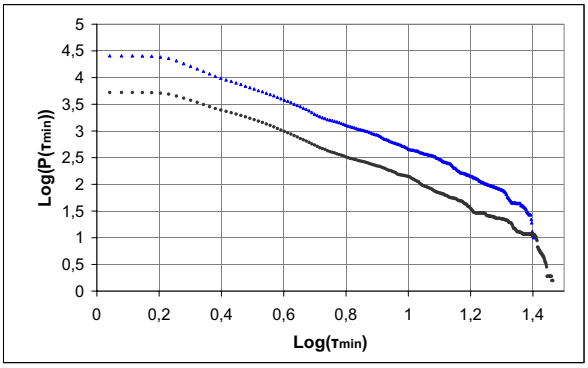

Fig. 4. Cumulative distributions of $\tau_{\min}$ in the PingER sample 2002, in log-log scale, for 100Byte- (upper curve) and 1000Byte-packets (lower curve). The lower curve has been shifted downwards by 0.5 for clarity.

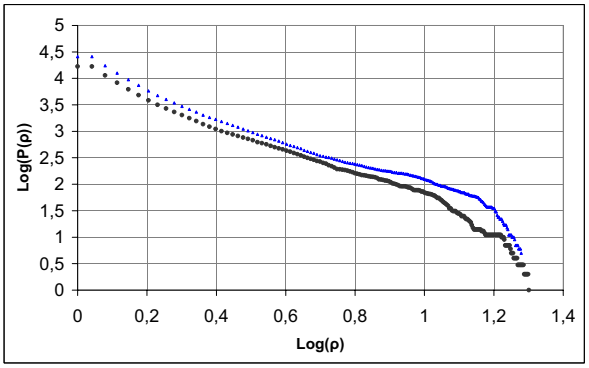

Fig. 6. Cumulative distributions of ρ in the PingER sample (2002), in log-log scale, for 100Byte- (upper curve) and 1000Byte-packets (lower curve). The vertical shift is due entirely to the smaller statistics of the 1000-Byte sample.

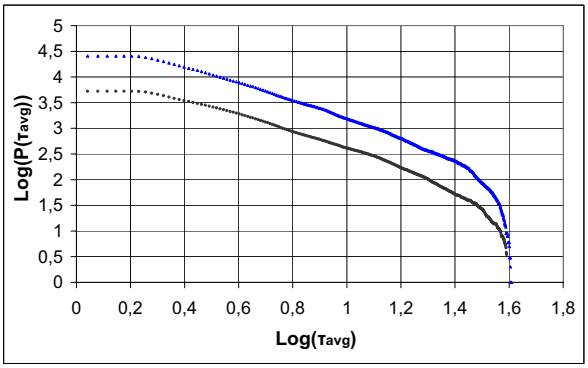

Fig. 5. Cumulative distributions of τ_{avg} in the PingER sample (2002), in log-log scale, for 100Byte- (upper curve) and 1000Byte-packets (lower curve). The lower curve has been shifted downwards by 0.5 for clarity.

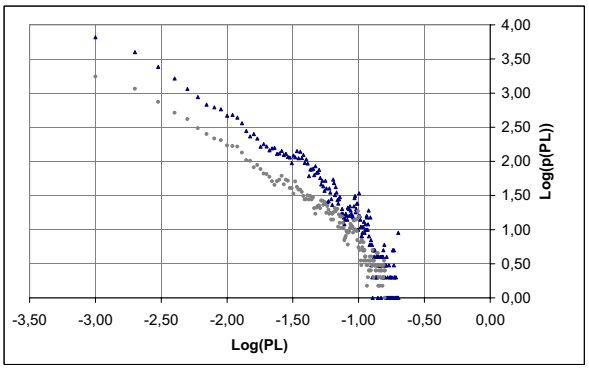

Fig. 7. Probability density distribution of Packet Loss in the PingER sample (2002), in log-log scale, for 100Byte- (upper curve) and 1000Byte-packets (lower curve). The lower curve has been shifted downwards by 0.3 for clarity.

Fig. 5 shows the cumulative distribution of τ_{avg}, again in log-log scale. The tail is slightly convex, but can still be approximated quite well by a straight line with slope -1.6, so α_{avg}=2.6 in this sample.

Fig. 6 shows the cumulative distribution $P(\rho)$. It is quite concave, implying a deviation from a pure power law. It has a much faster decay than the other two, with an exponent α_ρ close to 5 for first part of the curve (lying between ρ=1.1 and ρ=2.3). While these points cover only a fraction of the range of observed values of ρ, they comprise 92% of the whole sample and therefore are the most significant part of the distribution.

Finally, Fig. 7 shows the probability density distribution $p(PL)$. It has the shape of a broad bump, of which only the right side is shown. The region between PL=0.1% and PL=10% can be reasonably well approximated by a power-law with exponent α_{PL}=1.2.

Altogether, these 2002 results show a slight improvement relative to the ones discussed in [3], which referred to the years 2000-2001. To highlight this trend, we give in Table I the values of the best fits to the exponents, evaluated over each semester. (The values of α_ρ are the best fits from ρ=1.1 to ρ=7).

We recall [13], that the average RTT in the sample has been steadily decreasing at a rate of 10-20% per year, while PL has been improving at a rate of approximately 50% per year. In the case of the τ variables, this general improvement in performance does not only show up in the average values but also in the positions of the peaks of the distributions and in the exponents of the tails.

B. Gloperf data

In order to have an independent check of the existence of the power-law tails, in [3] the data from the Globus testbed were also analyzed. These data were collected in the period August-October 1999 and consisted of 17629 unique measurements between 3158 host pairs on 138 different hosts [5]. The sample

TABLE I

TIME EVOLUTION OF THE EXPONENTS IN THE PINGER SAMPLE

	I sem 2000	II sem 2000	I sem 2001	II sem 2001	I sem 2002	II sem 2002
$\alpha_{\min}$	2.58	2.49	2.68	2.84	3.08	3.26
$\alpha_{\rm avg}$	1.95	2.00	2.09	2.06	2.29	2.59
α_{ρ}	3.6	3.6	3.7	3.8	4.0	4.0
$\alpha_{\rm PL}$	1.04	1.08	1.28	1.24	1.23	1.35

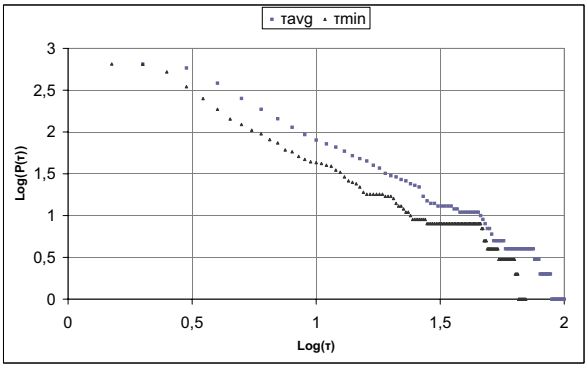

Fig. 8. Cumulative distribution of $\tau_{\min}$ (lower curve) and $\tau_{\rm avg}$ (upper curve) in the Gloperf sample, in log-log scale.

is thus both older and smaller than the PingER sample. The distribution of τ for all the hosts in the sample shows large irregularities. The reason is that several hosts were located at the same sites and this produces an artificial enrichment of data at specific site pairs. It is therefore necessary to consider site pairs, rather than host pairs. As a preliminary step, therefore, for each site pair we have calculated the minimum and the average RTT over all hosts belonging to those sites. This has reduced the sample to 650 unique site pairs. The distributions of $\tau_{\min}$ and $\tau_{\rm avg}$ for this sample are shown in Fig. 8.

The distributions are quite similar to the ones in Figs. 4 and 5. The peak of the distribution of $\tau_{\min}$ is located at $\tau_{\min}$=2.5. The exponent $\alpha_{\min}$ turns out to be equal to 2.7. Similarly, the peak for the distribution of $\tau_{\rm avg}$ is at $\tau_{\rm avg}$=3, and the exponent $\alpha_{\rm avg}$ is 2.5. There were no PL data for this sample. Altogether, the peaks of the distributions are significantly higher than the ones of the PingER data set and the exponents are also smaller in absolute value. Presumably this is due to the fact that the performance at the time, between the hosts under consideration, was indeed worse than the one for the PingER hosts. This is reasonable, given that the Gloperf data were collected from nine to six months earlier than the earliest PingER data that we have considered and given the trend of Table I. The main point, however is that the existence of the power law tails is confirmed and the exponents are found to be reasonably close to those of other data sets.

C. TTM data

The TTM (Test Traffic Measurement) is a measurement infrastructure designed and run by RIPE [9] as a commercial service offered to ISP's. It differs from the other infrastructures in that it measures one-way delays rather than RTTs. This makes it more useful both as a diagnostic and a scientific tool, but requires the use of GSM antennas in order to synchronize the clocks of the test boxes with a precision of a fraction of a millisecond. This makes the installation and maintenance of the test boxes more complicated.

¿From the point of view of a large-scale analysis like ours, the presence of synchronization errors could have an effect on the overall results. Hopefully, the influence of such errors will cancel out on average.

We have analyzed the data collected by TTM during the month of August 2002, at which time there were 48 active boxes. Each box sends one packet per minute to every other box. The values of $\mathrm{RTT}_{\min}$ and $\mathrm{RTT}_{\rm avg}$ for a given path are therefore the minimum and the average over two and a half millions of individual pings.

The distributions of $\tau_{\min}$, $\tau_{\rm avg}$ and ρ are shown in Figs. 9 and 10 respectively. The first two are very similar to the PingER distributions. The probability distribution of $\tau_{\min}$ has a peak at $\tau_{\min}$=1.8 and a tail with exponent $\alpha_{\min}$=2.9, whereas the probability distribution of $\tau_{\rm avg}$ has a peak at $\tau_{\rm avg}$=2.2 and a tail with exponent $\alpha_{\rm avg}$=2.6. Both have a clear cutoff for τ approximately equal to 30.

On the other hand, the distribution of ρ is quite concave for small ρ, then becomes convex without a clear cutoff. It is not very significant to approximate this distribution by a power-law. In the spirit of a Taylor expansion, one can approximate the curve in Fig. 10 by a straight line for small intervals of values of ρ; for example, in the region between ρ=1.1 (Log(ρ)=0.041) and ρ=2 (Log(ρ)=0.30), one can reasonably well approximate the distribution by a power-law with α_{ρ}=5.4. This is far too small a range of values to speak of a power-law tail, but still it is of some significance, insofar as this range contains 84% of all datapoints. The main point to be derived from this discussion is that the distribution of ρ has a fall-off that is much faster that that of the other two variables, and is also quantitatively very close to the one observed in the PingER sample.

D. AMP data

The AMP (Active Measurement Project) infrastructure consists of approximately 130 monitors collecting data, most of them at NSF-funded HPC-sites. See [8] for more detail. The monitors perform RTT, PL, topology and throughput tests. We have analyzed the RTT and PL data collected by AMP in the month of October 2002. After discarding invalid data, it has yielded a sample of 11265 paths for which the distance, the monthy minimum and average RTT and the monthly average PL were calculated.

The results are shown in Figs. 11-13. Fig. 11 shows the plots of the cumulative distributions of $\tau_{\min}$ and $\tau_{\rm avg}$. The cumulative distribution $P(\tau_{\min})$ has a plateau for $\tau_{\min} < 2.0$

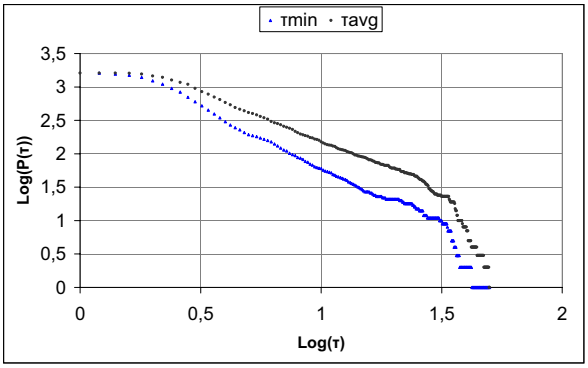

Fig. 9. Cumulative distributions of $\tau_{\min}$ (lower curve) and τ_{avg} (upper curve) for the TTM sample. The value of $P(0)$ is the total number of points in the sample and is the same for both curves.

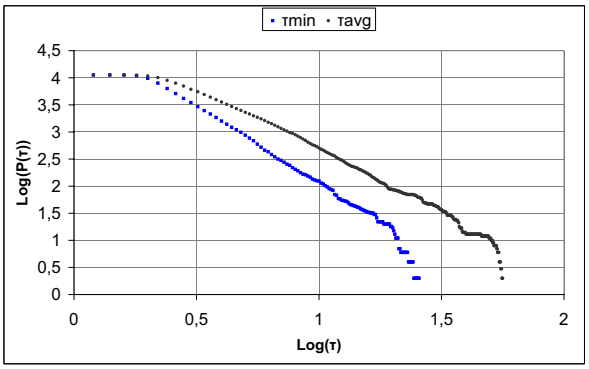

Fig. 11. Cumulative distribution of $\tau_{\min}$ (lower curve) and τ_{avg} (upper curve) in the AMP sample, in log-log scale. The two curves have the same values at $\tau_{\min}$=1, which is just the dimension of the sample.

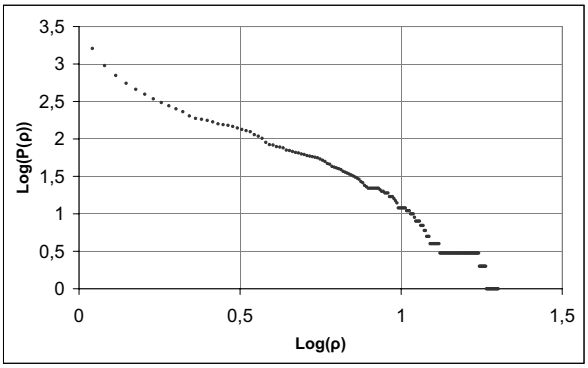

Fig. 10. Cumulative distribution of ρ for the TTM sample.

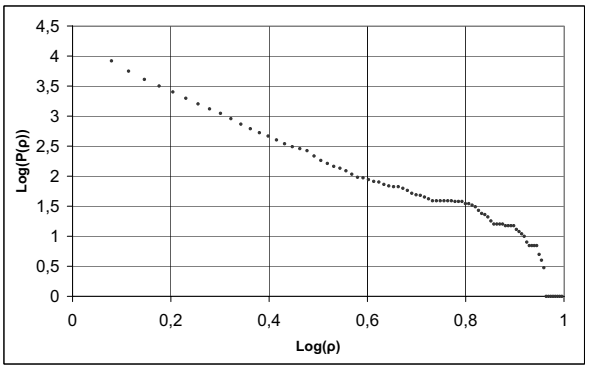

Fig. 12. Cumulative distributions of ρ in the AMP sample, in log-log scale.

(corresponding to the peak of $p(\tau_{\min})$) and shows a cutoff at $\tau_{\min}$=20. The region in between is very well approximated by a straight line with slope -2.8, yielding an exponent $\alpha_{\min}$=3.8. This result is significantly better than that of the other samples. The distribution of τ_{avg} has a tail that, between τ_{avg}=2.2 and a cutoff at τ_{avg}=47, is fairly well described by a power law, but shows a slight convexity. The best fit yields a value for α_{avg} of 3.3.

Fig. 12 shows the distribution of ρ. Like the one of the PingER sample (Fig. 6), it has a tail exhibiting a concavity in a log-log plot. However, the end of the tail, which curves upwards from linearity, contains very few points. If we restrict ourselves to 99% of the data points, the fit with a power law is very good and yields an exponent very close to 5.

The distribution of the packet loss, shown in Fig. 13, is a convex function with marked irregularities at packet losses of order 2-4%. These may be due to some of the monitoring hosts being out of order for some fraction of the time, thus artificially enriching the PL data at specific values. If we

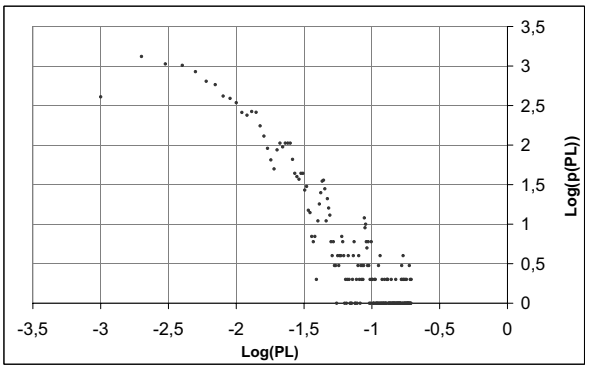

Fig. 13. Probability density distribution of Packet Loss in the AMP sample, in log-log scale.

TABLE II

SUMMARY OF RESULTS

Sample	Variable	Mode	α
PingER 100	$\tau_{\min}$	1.75	3.2
PingER 100	τ_{avg}	1.9	2.6
PingER 100	ρ	$<$1.1	$\approx$4.5
PingER 100	PL	-	1.2
PingER 1000	$\tau_{\min}$	1.8	3.1
PingER 1000	τ_{avg}	2.0	2.5
PingER 1000	ρ	$<$1.1	$\approx$4.5
Gloperf	$\tau_{\min}$	2.5	2.7
Gloperf	τ_{avg}	3	2.5
Gloperf	ρ	$<$1.2	$\approx$3.6
TTM	$\tau_{\min}$	1.8	2.9
TTM	τ_{avg}	2.2	2.6
TTM	ρ	$<$ 1.1	$\approx$5
AMP	$\tau_{\min}$	2.0	3.8
AMP	τ_{avg}	2.4	3.3
AMP	ρ	$<$1.1	$\approx$5
AMP	PL	-	1.5

TABLE III

NUMBER OF HOSTS PER AREA, IN EACH SAMPLE

	PingER	Gloperf	TTM	AMP
Europe	97	11	42	0
North America	85	29	4	106
South America	16	0	0	0
Asia-Pacific	5	4	2	1

smooth them out, it is possible to fit part of the curve, from the peak at PL=0.4% to roughly PL=2.5%, with a power law with exponent 1.5, in good agreement with the PingER data.

E. Summary of the data

Table II is a synopsis of the data. For each sample and each variable, we give the mode of the probability distribution $p(t)$ and the corresponding exponent that best fits the power-law tail.

The data for $\tau_{\min}$ and τ_{avg} show a remarkable agreement between PingER, Gloperf and TTM, with exponents close to 3 and 2.5 respectively, while AMP has significantly better distributions, with exponents 3.8 and 3.3. In order to discuss further these differences, it is instructive to look at the geographical distribution of the hosts involved in the four samples that we have examined, which is given in Table III.

The TTM and AMP samples are localized mostly in Europe and in North America, respectively, while the PingER and (to a lesser degree) Gloperf samples have a wider geographic reach and therefore are presumably a more representative sample of the Internet as a whole. (One has also to remember that the TTM test-boxes are located mostly on commercial networks, while the other samples consists mostly of academic sites.)

The difference between the Gloperf and PingER distributions, especially the higher mode but also the lower exponents, can probably be explained by observing that the Gloperf data were collected in 1999 while the PingER data shown here refer to the year 2002. The time evolution of the distributions emerging from the numbers shown in Table I is probably sufficient to explain the differences.

Differences in the behaviour of the samples may also reflect differences in the local network infrastructure. It is possible that the slightly lower value of $\alpha_{\min}$ (worse performance) shown by the TTM data relative to the other samples is due to the influences of geopolitics on the structure of networks in Europe.

It is not clear from these data whether the excellent behaviour of the AMP hosts reflects the general state of North American networks or is peculiar to this specific sample. One has to bear in mind that almost all the AMP probes are connected at high speed to national backbones such as Abilene. As we shall argue in Section IV, hosts connected to a single backbone have generally lower τ than hosts connected to different backbones.

A partial answer comes from analyzing the PingER data after separating the hosts according to their geographical area. The PingER 2002 data yield exponents $\alpha_{\min}$=2.5 for Europe (including Russia, the Balkans and Eastern Europe) and 2.7 for North America, confirming to some extent the expectations that North American networks are in some way better meshed that European ones. On the other hand, these values are markedly lower than the AMP exponents, suggesting that the AMP sample is indeed quite untypical.

Incidentally, the reason why the whole PingER sample shows better behaviour than the subsets of European and North American hosts is that the intercontinental paths are usually dominated by a single rather straight hop, and this tends to improve the performance. If we consider the probability distribution of paths consisting of pairs of hosts on opposite sides of the Atlantic, it has a peak at $\tau_{\min}$=2, a cutoff at $\tau_{\min}$=10 and a very fast fall-off with an exponent of order 4.8.

All these distributions exhibit an approximate power-law tail. In the case of Packet Loss, $\tau_{\min}$ and τ_{avg} this behaviour extends over one or two orders of magnitude while in the case of ρ, which has a much faster decay, it covers only a relatively small range of values of the variable. In the discussion of the TTM data we have already made some cautionary remarks on the significance of power-law tails with high exponents. In general, the relevance of power-laws is in the very different behavior from Poissonian or regular distributions. In the latter, it is possible to identify the average value of the distribution as a characteristic parameter of the system, since the majority of the events will be close to this value (the mean is very close to the mode of the distribution and departures from the mean are exponentially unlikely). In power-law behavior the average value is not typical: it does not express a particular point of the distribution. Usually many events will be smaller than the average value and there is an appreciable probability

to find large deviation from the average. This is mathematically signalled (if the exponent $\alpha \leq 3$) by the fact that power-law distributions have divergent fluctuations since the second moment is in principle unbounded. It is then the physical cut-off that determines the level of fluctuations, which is usually very large.

Finally, we remark that the Packet loss data are sometimes hard to interpret due to the fact that we cannot discriminate between network problems and host problems. For example, we have already remarked that some of the bumps that are observed in the distribution in Fig. 13 may be due to some of the hosts being out of order for some time. Since all other hosts will record these events as lost packets, this will produce an anomalously large number of host pairs with a specific percentage of PL. Of course, we are interested in the PL due to the network, so in future experiments it will be advisable to have a method to validate a lost packet as a network failure rather than a monitoring host failure.

IV. Simple network models.

Given that the quantity $\tau_{\min}$ is an estimate of the deviation of a path from being straight, we expect that its distribution has a geometrical origin. To this end, it is instructive to look at some simple models.

Consider first a uniform, random distribution of points on a plane, all connected to each other via a central hub. Without loss of generality we can assume at first that the hub is located in the origin. We are interested in the distribution of $\tau_{\min}$ for such a network. If we neglect processing delays, $\tau_{\min}$ is the ratio of the "network distance" ℓ to the geometric distance d. The length of a network path, ℓ, is given by the sum of the distances of the points to the hub. A high value of $\tau_{\min}$ implies that the distance between the two points is much smaller than either of their distances from the hub; therefore, in the limit, the distances of the points from the hub are approximately the same.

Let us first keep one of the endpoints fixed and let r be its distance from the hub. ℓ is approximately equal to $2r$. The number of points that can be reached by paths that have $\tau_{\min}$ larger than a given value t is equal to the number of points whose distance from the first, fixed endpoint is less than $2r/t$. With a uniform distribution, this is proportional to the area of the disk with radius $2r/t$ and hence scales like t^{-2}. The total cumulative distribution $P(\tau_{\min})$ is given by summing the partial distributions over all the endpoints; since all these distributions have a t^{-2} tail, so does the global distribution of $\tau_{\min}$. The probability $p(\tau_{\min})$ is obtained by differentiating the cumulative distribution. We have thus found that the density distribution of $\tau_{\min}$ in a planar, uniform, star-shaped network has a power law tail with exponent $\alpha_{\min}$=3. This is very close to the value observed in some of the samples. Still, other samples show other exponents.

One can try to change some of the assumptions to make the model more realistic. First, the assumption of a uniform distribution of points is certainly very unrealistic. It is reasonable to assume that the geographic distribution of routers follows the geographc distribution of the population (for a discussion of this point see [12]). This distribution is a fractal, with the number of points within a circle of radius $2r/t$ growing like $(2r/t)^{1.5}$ [16]. This would change (for the worse) the exponent $\alpha_{\min}$ from 3 to 2.5.

Another consideration is that the distribution cannot be infinite. If we assume a planar geometry we have to put a bound on r. A sharp cutoff on radius produces a distribution with a peak for small $\tau_{\min}$, since in such geometries there are many points that are almost aligned at opposite sides of the hub. Let us consider instead more realistic "gaussian" worlds consisting of points randomly distributed in the plane with a gaussian density distribution centered at the origin.

This geometry may be a reasonable approximation for a European network, for example, which has a finite size and where the effects of the curvature of the Earth can be neglected. This model violates the assumption of a uniform distribution made above, but still one expects a tail with the same exponent. In fact, as noted above, the tail consists of pairs of points that are very close to each other; if the density is sufficiently high, the average distance between points is much less than the typical distance over which the density varies (in the gaussian world, the variance) and therefore the tail is dominated by pairs of points for which the density may be approximated by a constant.

This is confirmed by numerical simulations. Fig. 14 shows the cumulative distributions $P(\tau_{\min})$ for various Gaussian worlds. When the hub is located in the origin, the distribution is almost entirely dominated by the tail, with $\alpha_{\min}$=3. Now suppose that the hub is moved away from the center. As it moves away, it becomes less likely for a pair of points to be aligned at opposite sides of the hub and therefore the peak of the density distribution $p(\tau_{\min})$ moves towards higher values and the cumulative distribution is shifted upwards. However, the tail has always the same exponent $\alpha_{\min}$=3. Even when the hub lies entirely outside the Gaussian "cloud", this argument remains valid.

One way to improve network performance would be to have more than one hub. Suppose there are several hubs, randomly distributed with the same Gaussian distribution as the points, but with lower density. Assume that two points communicate via the nearest hub (more precisely, for a given pair of points, the "network distance" is given by considering the sums of the distances from the given points to each hub, and then taking the minimum of this set). Could this change the exponent? Not really. Since the distribution of points is much denser than the distribution of hubs, the average distance between hubs, $< d_H >$, is much larger than the average distance between points $< d >$. The highest values of $\tau_{\min}$ arise when the points are close to each other and as far from the hub as is possible. Typically, this will mean less than $< d_H >$. In this situation the closest hub is the only one that matters, so the tail will be the same as in the single hub case. Of course, high values of $\tau_{\min}$ are much less likely, so the distribution will have a peak closer to one and the tail will be depressed relative to the single hub case. This is confirmed by the results

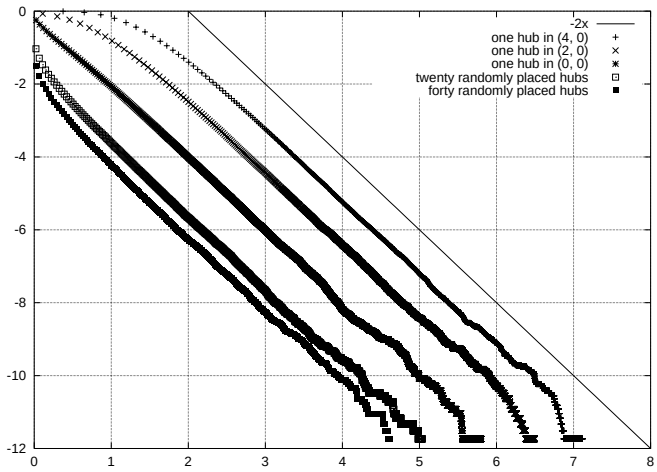

Fig. 14. Cumulative distributions of $\tau_{\min}$ in Gaussian worlds consisting of 1000 points, for various positions of a single hub and for various random distributions of hubs.

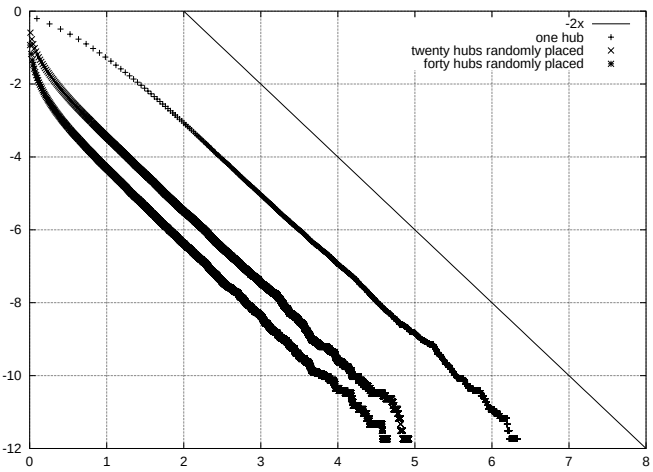

Fig. 15. Cumulative distributions of $\tau_{\min}$ in Spherical worlds for various random distributions of hubs.

of simulations with 20 or 40 hubs, which are the two lowest curves in Fig. 14.

Another effect that will play a role in shaping the distribution of $\tau_{\min}$, especially on a global scale, is the spherical geometry. This automatically puts a cutoff on distances and is also more realistic if we want to consider the whole Internet. Consider first the case of a single hub. Since the distribution for large $\tau_{\min}$ is determined by local behaviour, we expect again a tail with $\alpha_{\min}$=3. Small $\tau_{\min}$ corresponds to pairs of points that are almost aligned on opposite sides of the hub. In the case of a uniform distribution in a plane, the number of such pairs grows linearly with distance from the hub. By contrast, on a sphere the number of such pairs only grows for distances up to $\pi/2$ and then decreases again. One therefore expects fewer points with small $\tau_{\min}$. This is confirmed by Fig. 15, which shows the results of several simulation for a uniform distribution of points on a sphere of unit radius. It gives the distributions for one, twenty or forty hubs. In all cases the distributions have the $\tau_{\min}^{-3}$ tails and the effect of increasing the number of hubs is only to shift the peak to lower $\tau_{\min}$, without changing the exponent of the tail.

Comparing the single-hub distributions in the Gaussian and Spherical worlds (middle line in Fig. 14 vs. top line in Fig. 15) we see that the former is essentially straight, while the latter is slightly convex for small $\tau_{\min}$. (A uniform distribution with a sharp cutoff on radius would produce a concave distribution for small $\tau_{\min}$.) A similar comparison in the case of multiple hubs shows that the distributions become more and more the same as the number of hubs increases. This was to be expected, since in the presence of many hubs it is the local geometry that matters and the sphere looks more and more like a plane when one goes to smaller scales.

The lesson of these simulations is that the power law tail with exponent $\alpha_{\min}$=3 is a very robust feature of hub-dominated networks. In order to change the exponent, one has to introduce some degree of meshedness in the network. Consider another model where the points communicate through a transport network, consisting again of randomly distributed Points of Presence (POPs), having again a gaussian probability distribution but with different normalizations (there are many more points than POPs). The POPs form a fully meshed network, and the points communicate with each other through the network, accessing it at the nearest POP. The network distance between two points is therefore the sum of the distances from the two points to the nearest POPs, plus the distance between the POPs.

One expects this model to exhibit a better behaviour than the previous ones. The results of the simulations for a Gaussian world of this type are shown in Figs. 16 and 17. The number of points is always equal to 1000, as before. In Fig. 16 the Gaussian distribution of the POPs has the same width as the Gaussian distribution of the points and the number of POPs is varied. The case with one POP coincides with the Gaussian world with a single hub. As the number of POPs is increased, the probability distribution develops a marked peak near $\tau_{\min}$=1 and the exponent of the tail also changes. For a density of POPs equal to one tenth the density of points, $\alpha_{\min} \approx 3.4$.

In Fig. 17 the number of POPs is held fixed (at 1 POP every 40 points) and the width of the Gaussian distribution of the POPs is varied. When the width of the distribution of POPs (RH) becomes small relative to the width of the distribution of points (RN) the distribution tends to the one for a single hub. As the RH increases, the distributions tend to the ones given in Fig. 16.

V. Comments and conclusions

The main point of this paper is that the quality of an IP network should be measured using dimensionless metrics. A number of such metrics can be extracted from the distributions of PL (which is already dimensionless) and of the ratio $\tau = \mathrm{RTT}/d$. All these distributions have a single maximum and decay to zero with varying speed. For some of them this decay is very well approximated by a power-law.

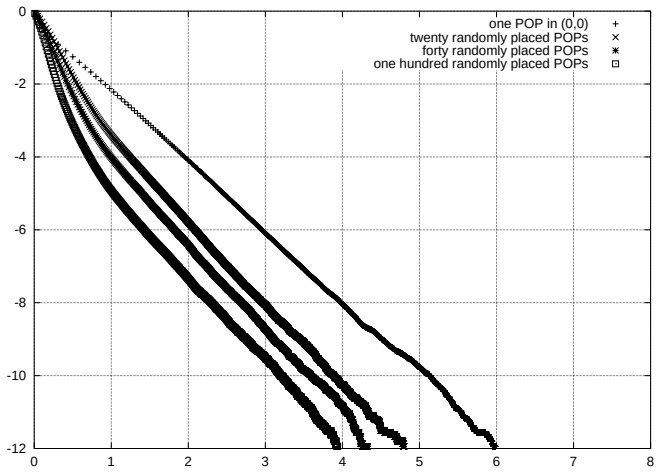

Fig. 16. Cumulative distributions of $\tau_{\min}$ in Gaussiam worlds. 1000 points communicate through a perfectly meshed transport network, having a Gaussian distibution of POPs. The density of POPs is varied.

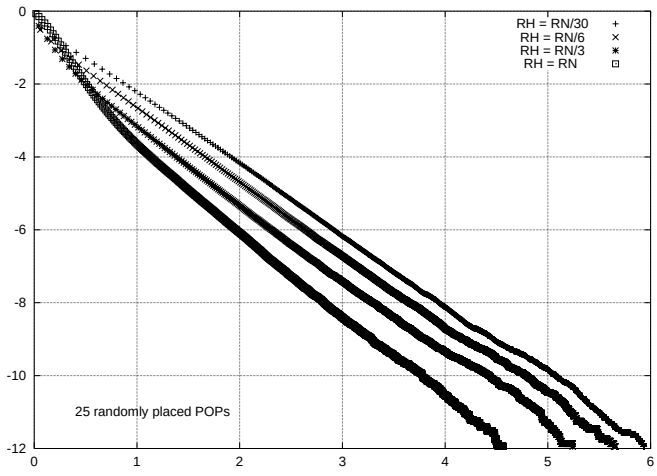

Fig. 17. Cumulative distributions of $\tau_{\min}$ in Gaussiam worlds. 1000 points communicate through a perfectly meshed transport network, having a Gaussian distibution of POPs. The radius RH of the distribution of POPs is varied.

The distribution of $\tau_{\min}$ shows in all samples a very clear power-law tail with $\alpha_{\min}$ close to 3. We believe this tail, and the exponent, to be a genuine feature of the Internet. Subnetworks of the Internet or private IP networks may well exhibit different exponents, as exemplified by the AMP sample. Simple (though unrealistic) network models predict a tail with $\alpha_{\min}$ equal or close to 3, which is surprisingly close to the observed value.

The distribution of τ_{avg} also has a power-law tail, but the fit is not so good in all samples: some tend to exhibit a slight convexity (in log-log scale). The distribution of ρ has a peak which is very close to one and has a very fast decay with α_ρ of the order of 4-5. The power-law approximation is not so good for this variable. The distribution of PL has the shape of a wide bump (in log-log graph). The tail is convex. The region between 0.1% and 10% packet loss is approximated quite well by a power-law, with $\alpha \approx 1.2$.

We have considered some of the factors that have a role in shaping the distribution of $\tau_{\min}$. Hub-dominated networks, either with one or several hubs, and a uniform density of points, have a power-law tail with exponent $\alpha_{\min} = 3$. A lower exponent (worse performance) will arise in the case of a fractal distribution of points, which is known to be a more accurate model for the density of hosts on the Internet. A higher exponent (better performance) will arise when the network has some degree of meshedness. A slight convexity will also arise when the spherical geometry of the earth is taken into account.

In trying to compare these very simple models to the real Internet, there are several issues that have to be addressed.

First of all, is it really meaningful to talk of such things as "the distribution of $\tau_{\min}$ for the Internet as a whole"? One has to make many arbitrary choices about what hosts to count as being part of the Internet and what not (*e.g.*, are dialup lines to be counted?). In order for "the distribution of $\tau_{\min}$ for the Internet as a whole" to be a meaningful concept, one would have to prove that it is relatively insensitive to these choices. If it is not, then one should only talk of "the distribution of $\tau_{\min}$ for this set of hosts".

Another question that we have entirely avoided is: how much of $\tau_{\min}$ is due to the IP layer and how much to the underlying Layer two links? Since optical cables have a cost that is proportional to their length, we strongly supect that layer two links have a much better distribution of ℓ/d. If this is true, then the observed long tails in the distribution of $\tau_{\min}$ and τ_{avg} are due to the economics of IP – essentially the fact that Internet transit is billed in a distance-independent fashion (a fact that has come to be known as the "death of distance" [14]).

The Internet is divided into a few tens of thousands of administrative domains called Autonomous Systems (AS's), each corresponding roughly to an ISP or a relatively large organization. If we give for granted that layer-two links are almost straight, and that the observed degree of wiggliness of the paths is due mostly to IP, we can further ask if there is a significant difference between intra-AS and inter-AS paths. There are strong hints that intra-AS paths tend to be far better than inter-AS paths. This is supported by analyses of some carrier backbones [15] as well as the work of [6]. Large commercial networks appear to be quite meshed at the national/continental level and therefore network paths tend not to be unduly long. On the other hand in most countries inter-AS peerings tend to be concentrated at one or only a few large peering points, which then effectively act as hubs. Such situations may be modelled reasonably well by one of the Gaussian worlds described in Section IV. As observed in Section I, this is also a situation that can easily give rise to very high values of $\tau_{\min}$. The only way to improve these situations would be to improve the mesh by greatly increasing the number of regional and local peering points.

A hub-and-spokes topology may also prove a resonable approximation in modelling some intercontinental paths. As is well-known, in the early days of the Internet the US acted as

a giant hub, with much inter-European traffic travelling across the Atlantic. It is still true that communications between South America, the Asia-Pacific and Europe pass mostly through the US. It is only very gradually that a tighter mesh is arising.

Until now, network models builders [17], [18] have focused mainly on reproducing the distribution of the degree of connectivity of the real networks. We believe that it is useful to reproduce not just the topology but also the geometry of the networks. In this paper we have defined the geometric parameter $\tau_{\min}$ and we have looked at some of the factors that determine its distribution. It is possible that a suitable model that takes into account the general features of ASs, their internal and external connectivity, may be able to reproduce to a satisfactory degree all the features of the observed distributions of $\tau_{\min}$.

More realistic, dynamic network models will have to take into account also the features of the network traffic. We have not looked into this issue, but we claim that the observed distributions of τ_{avg} and ρ will be useful guides. In this connection we emphasize again that the fast decay of the distribution of ρ implies that the observed average degree of congestion is quite low. One has to remember however that these are monthly average values, comprising periods of intense network use and periods when the network is almost idle. A snapshot of the average of ρ over few hours during a weekday is likely to exhibit a much worse distribution.

ACKNOWLEDGMENTS

This work was carried out within the IPM project [11] of the Istituto Nazionale di Fisica Nucleare. A.V. is partially supported by supported by the European commission FET Open project COSIN IST-2001-33555.

We would like to thank Les Cottrell, Craig Lee, Henk Uijterwaal and Tony McGregor for access to their raw data and for useful correspondence and conversations.

REFERENCES

[1] G. Hooghiemstra, and P. Van Mieghem, “Delay distributions on fixed Internet paths”, Delft University of Technology report 2001 1031

[2] C.J. Bovy, H.T. Mertodimedjo, G. Hooghiemstra, H. Uijterwaal and P. Van Mieghem, “Analysis of End-to end delay Measurements in Internet”, Proceedings of the PAM 2002 Conference, Fort Collins, Colo, March 25-26 2002

[3] R. Percacci and A. Vespignani, “Anomalous fluctuations of the Internet global performance”, cond-mat/0209619;

[4] B. Huffaker, M. Fomenkov, D. Moore and kc claffy, “Macroscopic analyses of the infrastructure: measurement and visualization of Internet connectivity and performance”, Proceedings of the PAM 2001 Conference, Amsterdam, The Netherlands, April 23-24 2001

[5] C. Lee and J. Stepanek, “On future global grid communication performance”, 10th IEEE heterogeneous computing Workshop, San Francisco, California, May 2001

[6] L. Subramanian, V.N. Padmanabhan and R.H. Katz, “Geographic properties of Internet Routing”, USENIX 2002,

[7] PingER web site: `http://www-iepm.slac.stanford.edu/pinger/`

[8] AMP web site: `http://watt.nlanr.net/`

[9] TTM web site: `http://www.ripe.net/ttm/`

[10] IEPM web site: `http://www-iepm.slac.stanford.edu`

[11] IPM web site: `http://ipm.mib.infn.it/`

[12] A. Lakhina, J.W. Byers, M. Crovella and I. Matta, “On the Geographic Location of Internet Resources”,

[13] L. Cottrell, “Internet End-to-end Monitoring Project - Overview”, presentation for Internet 2 PiPES/SLAC IEPM meeting Jan 2003, reported at `http://www.slac.stanford.edu/grp/scs/net/talk/pipes-jan03.ppt`

[14] F. Cairncross, “The death of distance 2.0”, Harvard Business School Publishing, Cambridge (MA) 2002

[15] R. Percacci, H. Sarmadi, unpublished

[16] S.-H. Yook and H. Jeong and A.-L. Barabási, “Modeling the Internet’s large-scale topology”, Proc. Nat. Acad. Sci. USA, 99, 13382, 2002

[17] Dorogovtsev, S. N. and Mendes, J. F. F., “Evolution of networks”, Adv. Phys. **51**, 1079, 2002,

[18] Albert, R. and Barabási, A.-L., “Statistical mechanics of complex networks”, Rev. Mod. Phys.,**74**, 47, 2002