Brief Announcement

On the Internet Delay Space Dimensionality

Bruno D. Abrahao* Robert D. Kleinberg*
Dept. of Computer Science, Cornell University
Ithaca, NY 14850, USA
{abrahao,rdk}@cs.cornell.edu

Categories and Subject Descriptors: C.2.4 [Computer-Communication Networks]: Distributed Systems

General Terms: Measurement.

Keywords: Delay Space, Internet Structure.

1. INTRODUCTION

In this work we investigate the dimensionality properties of the Internet delay space, i.e., the matrix of measured round-trip times between Internet hosts. Properties of this space affect the performance of coordinate-based positioning systems. Previous work on network coordinates has indicated that the delay space can be embedded, with reasonably low distortion, into a low-dimensional Euclidean space. Our work addresses the following questions: to what extent is the dimensionality an *intrinsic* property of the delay space, defined without reference to a host metric, such as Euclidean space? Is the intrinsic dimensionality of the Internet delay space close to the dimension determined using embedding techniques? If not, what explains the discrepancy? What properties of the network contribute to its overall dimensionality?

We investigated all-pairs round-trip time (RTT) measurements between Internet hosts, collected via the King method. A fuller exposition of our research, including quantitative results omitted here for space reasons, will appear in [1].

2. DIMENSIONALITY MEASURES

There are many ways to define the dimensionality of a delay matrix. One could embed the points in $1, 2, 3, \ldots$ dimensions using an embedding algorithm such as Vivaldi [3], stopping at the lowest dimension D which permits an embedding with small quartiles of relative errors and declaring this dimension D to be the *embedding dimension*. The estimation of the embedding dimension of our datasets using Vivaldi indicates a value in the range from 4 to 7. Alternatively, principal component analysis (PCA) is frequently applied, but often provides only an imprecise dimensionality estimate, as in the case of our datasets.

In this work, we propose measuring the delay space using dimensionality notions, such as the *correlation* and *Hausdorff* dimensions, based on power laws in the distance data. Point sets whose distances satisfy power laws (often with non-integer exponents) are said to behave like *fractals* [2]. Accordingly, we refer to these metrics as fractal measures. To illustrate how these measures work, the correlation dimension can be computed by plotting in logscale the number of pairs of nodes that are within a given distance from each other. Figure 1 shows the plot generated from one of our datasets. A striking feature of the delay distribution is a power law that persists roughly over two orders of magnitude, i.e., from 3ms to 100ms. (Note that this range of latencies includes almost every Internet route that is not trans-oceanic.) Compared to the embedding dimension, the fractal measures revealed a much smaller dimensionality, less than 2 in all cases we observed. Similar results (not shown here, but reported in [1]) were obtained when using the Hausdorff dimension, which is the power-law exponent that emerges when counting the number of radius-r balls required to cover the space, as r tends to zero.

*Both authors are supported by NSF Award CCF-0643934.

PODC'08, August 18–21, 2008, Toronto, Ontario, Canada.
ACM 978-1-59593-989-0/08/08.

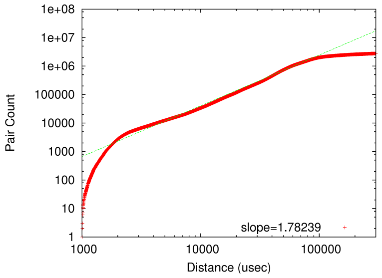

Figure 1: Pair-count plot of delays between 2385 Internet hosts. The correlation dimension, represented by the slope of the line, is 1.782.

3. ON THE DELAY SPACE STRUCTURE

The fractal measures reveal a dimensionality shift attributable to the structural configuration of the Internet's autonomous system (AS) topology, which is not captured by the embedding and PCA methods. This phenomenon can be shown by decomposing the network into overlapping Tier-1 AS networks — each composed of a Tier-1 AS and its downstream customers — and analyzing the geometry of each piece in isolation. We measured these subnetworks using the correlation dimension, Hausdorff dimension, embedding dimension, and PCA. All of the 8 subnetworks, with one exception, exhibit correlation and Hausdorff dimensionality around 10% *smaller* than that of the whole matrix. A similar reduction is not achieved by other kinds of decomposition, for instance, by decomposing the network into pieces of smaller *cardinality*, or of smaller *diameter*. Interestingly, delay space dimensionality reduction cannot be explained by a corresponding reduction using the geographic distances.

4. REFERENCES

[1] B. Abrahao and R. Kleinberg. On the Internet delay space dimensionality, 2008. in preparation.

[2] A. Belussi and C. Faloutsos. Estimating the selectivity of spatial queries using the 'correlation' fractal dimension. In *Proc. of VLDB*, 1995.

[3] F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi: a decentralized network coordinate system. In *Proc. of ACM SIGCOMM*, 2004.