A Meaningful Metric for IPv4 Addresses

Leigh Metcalf

lbmetcalf@cert.org

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

This material is based upon work funded and supported by Department of Homeland Security under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense.

IPv4 Metric Goal

We want a metric on IPv4 space that accomplishes the following:

1. Respects Routing Boundaries: A /8 is the largest allocation possible, so IP addresses in different /8s should be far apart.
2. Does not require outside information.
3. Temporally stable.
4. Easy to compute.

IPv4 as Integer

Metric: Translate each IPv4 address into the 32 bit integer it represents, and use absolute value.

The distance between 1.2.3.4 and 1.2.3.5 is one, which is reasonable since they are adjacent

Problem: So is the distance between 1.255.255.255 and 2.0.0.0

This metric doesn’t respect routing boundaries.

IPv4 as Routed Entity

A discrete metric, if two IPv4 addresses are routed by the same ASN the distance is 0, otherwise it is 1.

This method requires outside information and is not temporally stable.

IPv4 by Geographic Location

Using a GeoIP database, geographically locate each IPv4 and find the great circle distance between them.

Requires outside information, not temporally stable, and most of all…

It has been shown that GeoIP is fairly precise on the country level. Beyond that… not so much.

Our metric would rely on imprecise data derived by unknown means.

IPv4 as $\mathbb{R}^4$

We treat a.b.c.d as the vector $(a\ b\ c\ d)^T$ and use the Euclidean Metric on it.

The distance between 1.2.3.4 1.2.3.5 is 1, which is reasonable since they are adjacent.

However, so is the distance between 1.2.3.4 and 2.2.3.4

This metric doesn't respect routing boundaries

IPv4 as $\mathbb{R}^4$ Take 2.

We still treat a.b.c.d as the vector $(a\ b\ c\ d)^T$

However, the distance is now a weighted Euclidean metric.

Let a.b.c.d and w.x.y.z be IP addresses. Our metric is given by:

$$\sqrt{2^{24}(a-w)^2+2^{16}(b-x)^2+2^8(c-y)^2+(d-z)^2}$$

IPv4 as $\mathbb{R}^4$ Take 2.

The weights were chosen from the definition of CIDR. Each weight represents the number of IPv4 addresses in that network.

For example, a /8 represents 2^{24} IP addresses, so that was the weight for the first quad.

This is the weighted metric on IPv4 addresses

Weighted Metric on IPv4 Addresses

- $d(1.2.3.4, 1.2.3.5) = 1$
- $d(1.2.3.4, 2.2.3.4) = 2^{12}$

1. Respects routing, in particular /8s
2. Does not require external information
3. Temporally stable.

Applications of the Metric

I originally created this metric as a way to find ‘inside’ and ‘outside’ for network flow.

To test this, I use the LBNL data set. I chose a random pcap from the data set and extracted 870 IPv4 addresses were extracted from it.

I used medoids to cluster the IPv4 addresses. The optimum number of clusters was determined to be 15.

Applications of the Metric

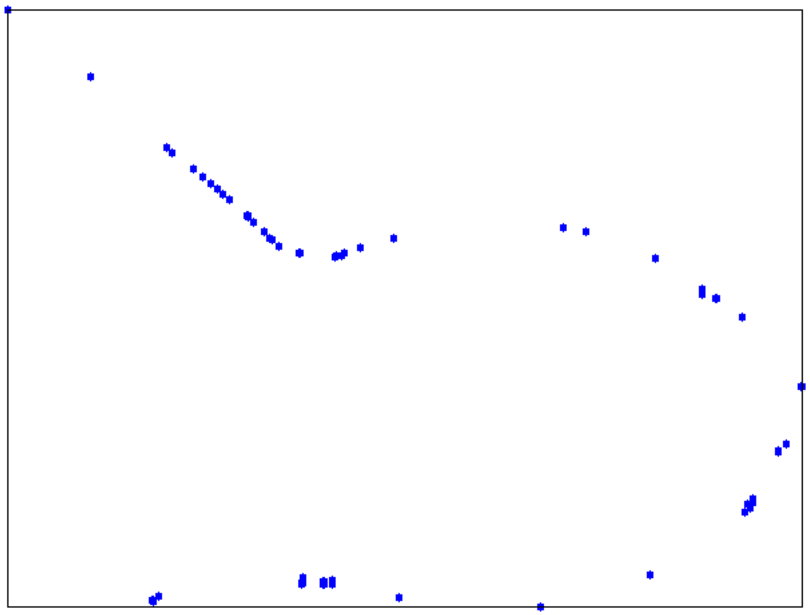
IP address	Size	IP address	Size
131.243.93.124	427	224.0.0.13	9
128.3.162.146	279	148.165.70.148	9
56.96.15.203	53	32.28.79.213	5
198.129.90.114	23	33.246.149.89	4
204.116.100.64	14	167.130.77.99	4
59.219.149.74	13	87.221.134.191	3
216.154.130.141	13	239.255.255.253	3
118.141.176.166	12		

Applications of the Metric

The two largest clusters were owned by the LBNL.

The second largest is owned by the Post Office

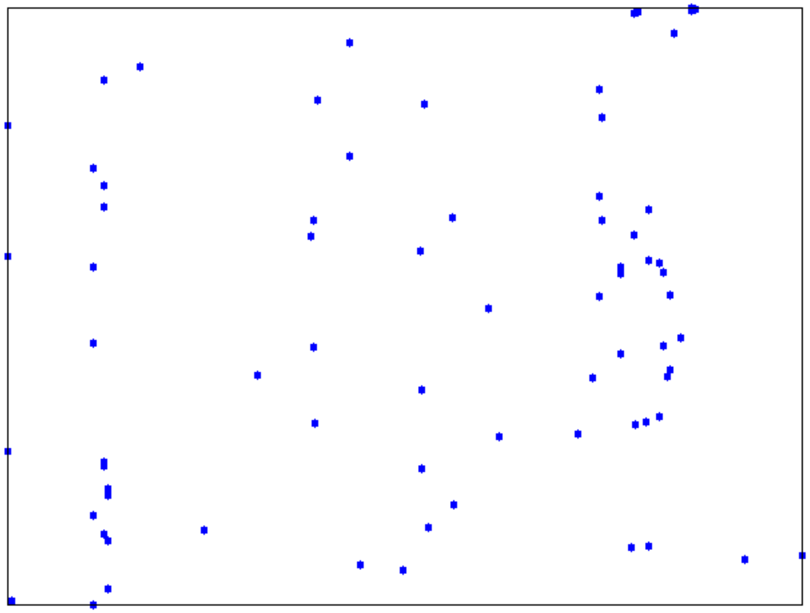
Applications of the Metric -- Visualization



Sammon’s nonlinear mapping for dimensionality reduction

Preserves relative distance between the higher dimension and the lower dimension

Applications of the Metric -- Visualization



PCoA – Principal Coordinate Analysis

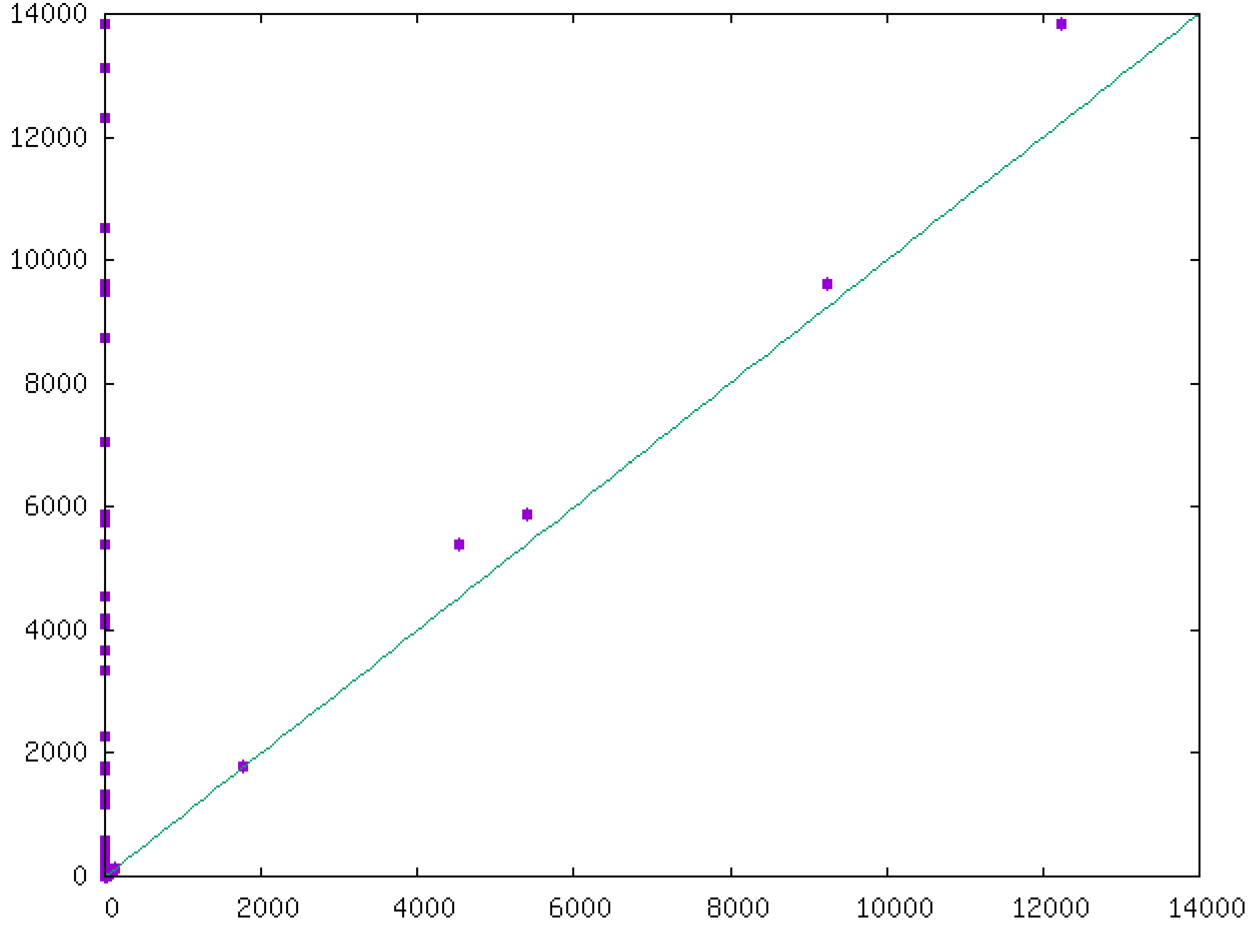
Also attempts to place things in lower dimensional space while retaining distance relationships.

Linear Method.

Applications of the Metric – Persistent Homology

Persistent Homology is the study of the shape of data.

Applications of the Metric – Persistent Homology



Questions/Comments?