Eyeball ASes: From Geography to Connectivity

Amir Rasti, Nazanin Magharei, Reza Rejaie, Walter Willinger

IMC 2010, Melbourne, Australia

Introduction

- Large body of research on the AS-level topology
  - Fueled by BGP- or traceroute-based measurement
  - The topology is modeled as a graph, AS as a node
- Concerns about the "graph view" of AS-level topology
  - Limited accuracy
  - Ignoring economics
  - Ignoring geographical coverage of an AS
- Geographic footprint of an AS affects its connectivity
  - e.g. AS X peers with Y if Y has certain on geo coverage, or y has certain number of overlapping PoP locations/IXPs with X.
- *How can we estimate geo- and PoP-footprint of an AS?*

This Paper

- Proposes a new approach to estimate ***geo-footprint*** and ***PoP-footprint*** of eyeball ASes
- Our approach complements traditional approaches
  - Relying on geo location of end-users
  - More accurate at the edge of the network, eyeball Ases
- Contributions:
  - A new approach to determine geo-footprint of eyeball Ases
  - Using geo-footprint to estimate and validate PoP-footprint
  - Leveraging inferred PoP locations & given AS-level topology to show that peering relationship at the edge is complex

Our Approach: An Overview

Four steps:

- Sampling end-users - IP address of Internet users
- Mapping end-users to geo locations
- Grouping end-users by AS using BGP information
- Estimating AS geo-footprint from location of its end-users

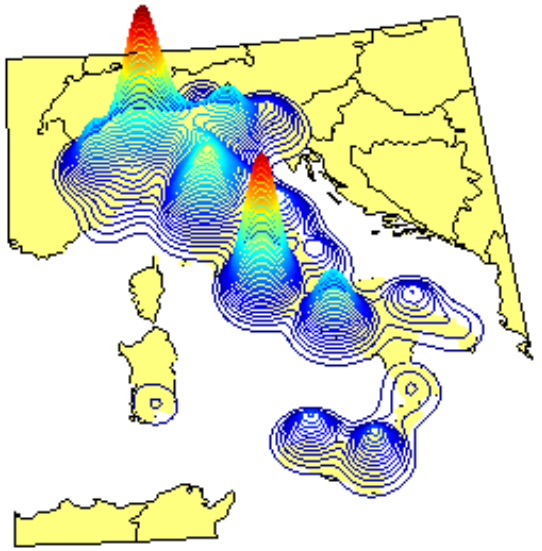
Sampling end-users

- Crawling popular P2P apps: Kad, BitTorrent, Gnutella
- IP to Geo mapping
  - Using GeoIP City & IP2Location
- Data Conditioning => target dataset
  - Removing IP address with large error (> 100km)
  - Removing ASes with less than 1K samples

	#Peers by source (M)			#ASes by level		
Region	Kad	Gnu	BT	City	State	Country
NA	1.2	8.9	1.7	36	162	129
EU	18	2.5	2.5	60	76	292
AS	17.8	1.6	1.0	117	35	134

Estimating Geo Footprint

- Using Kernel Density Estimation (KDE) method with Gaussian kernel function => probability density function
- KDE presents a weighted average across close-by peers
  - Smooth out the error in IP-geo mapping of individual users
  - Offers a more aggregate than user-level view
- Largest contour of the density function represents geo footprint
  - May consist of one or multiple regions

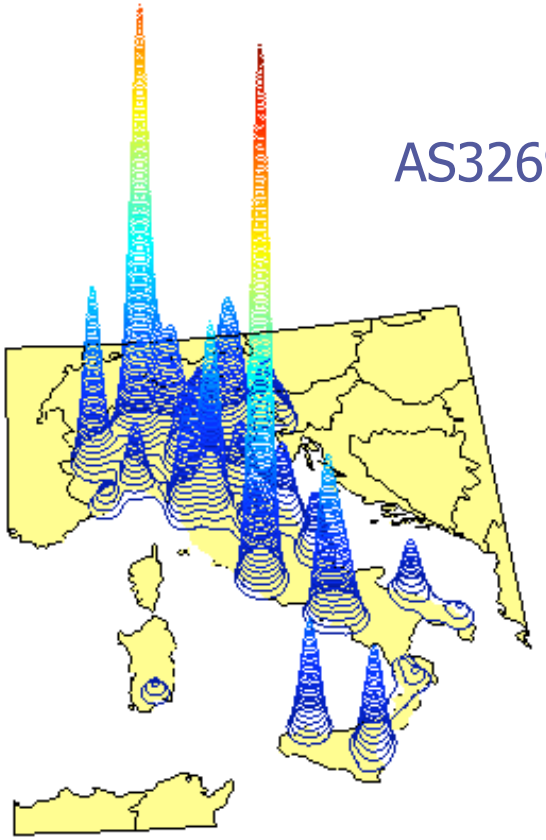


Setting Kernel Bandwidth

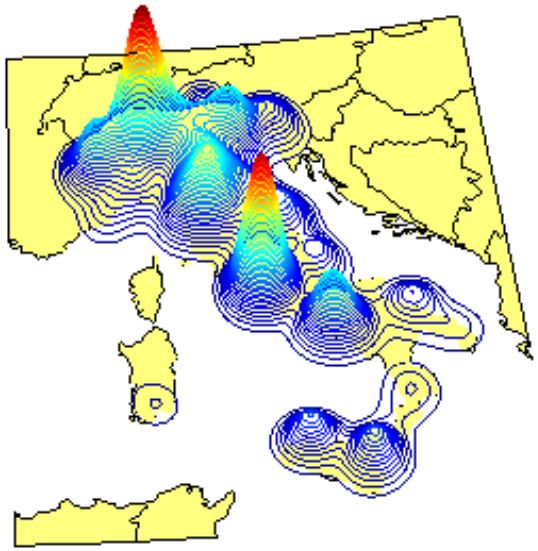
- Kernel bandwidth (BW) determines the scope of averaging
  - Larger BW filters out larger error but leads to a coarser resolution of geo-footprint
  - Accuracy of IP-geo mapping determines min bw for KDE
- We focus on ***city-level*** resolution for geo-footprint
  - Set kernel bandwidth to radius of a city: 40Km
- City level resolution reveals PoP locations

Effect of kernel Bandwidth

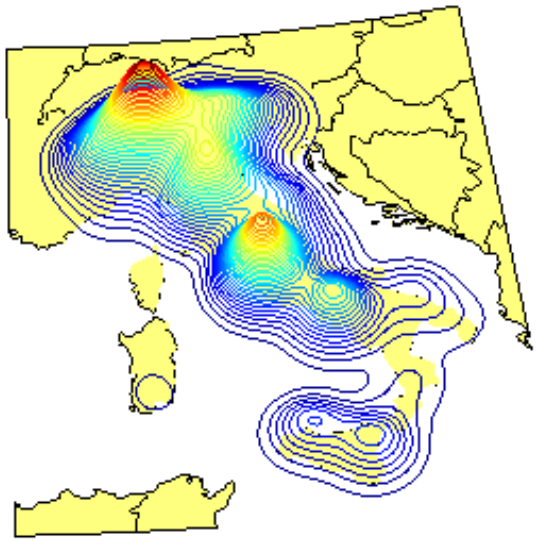
AS3269 : Telecom-Italia with 2.2 M samples



BW = 20km



BW = 40km



BW = 60km

PoP-Level Footprint

- Major cities in a geo-footprint with the highest user density are likely location of PoPs
- Identify coordinates of major local peaks of the density function
  - Filter out minor peaks due to randomly clustered samples
- Map each peak to the most populated city within the radius of BW from peak's coordinates (loose mapping)
- PoP level footprint is a collection of cities and their user density
- Biased samples may affect PoP-level footprint (see the paper for details)

Bias in Collected samples

- The fraction of collected samples from a city could be disproportional with actual user population per AS
- Cannot distinguish between market share of an AS in a city and penetration of P2P app in that city
- Mild bias only affects the density of identified PoPs
- Significant bias is unlikely with a large number of samples

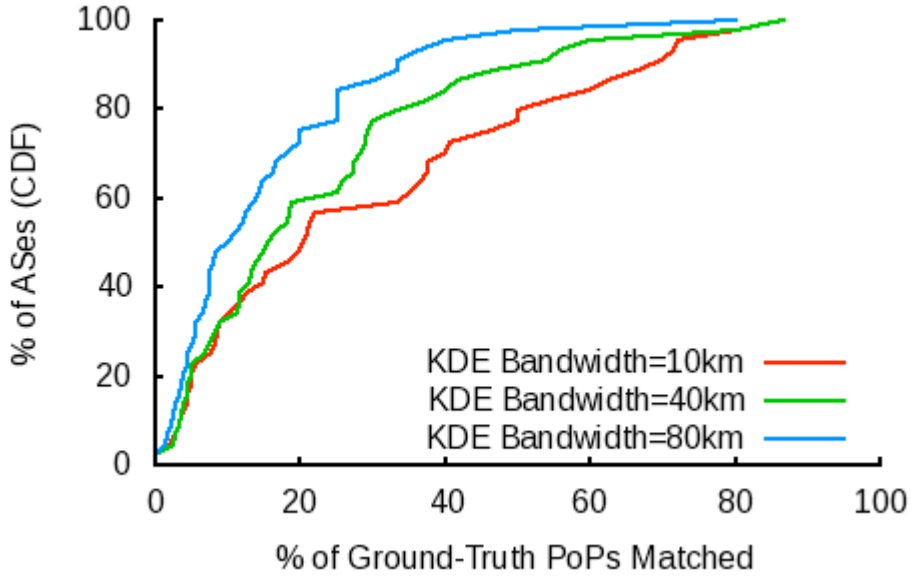
Evaluation

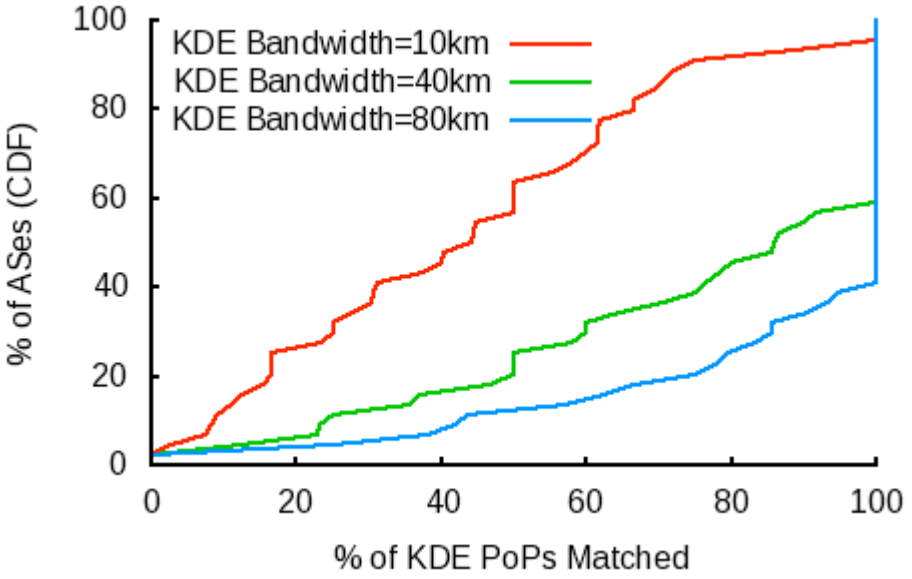
- Collecting reported PoP information for eyeball ASes on the Web as ground truth
  - Only available for a small fraction of ASes
  - Inconsistent terminology and method across different ASes
- Reference Dataset
  - Focused on 672 ASes
  - Identified PoP info for only 45 ASes (10 state-, 33 country-, 2 continent-level)
- Our approach identified 31.9, 13.6 and 7.3 PoPs/AS on avg as we increase BW 10km, 40km, and 80km, respectively.
  - Avg PoP/AS from ground truth is 43.7

Results

- Perc. of reference PoPs matched
- bw=40km, for the bottom 60% of ASes, < 20% of PoPs matched.
- Decreasing bw increases matched PoPs

- Perc. of KDE PoPs matched
- bw=80km, perfect catch for 60% of ASes, < 20% of PoPs matched.
- Decreasing bw reduces matched PoPs





- Using larger kernel bw leads to a smaller but more reliable set of PoPs for most ASes

Evaluation (cont'd)

- Possible sources of error against the ground truth
  - Some eyeballs ASes have PoPs away from their customers to connect to providers or peers
  - Some eyeball ASes have multiple close-by PoPs
  - Misinterpreted or obsolete info from the Web
- Please see the paper for
  - Comparison with PoPs detected by DIMES project
  - Case study that use geo properties

Summary

- The proposed approach is promising in identifying geo- and PoP-footprint of eyeball ASes
- Our case study demonstrates how geo information can be used to examine AS topology

Future Work

- Addressing the limitations of the technique
- Leveraging geo properties of ASes to examine their strategies to inter-connect