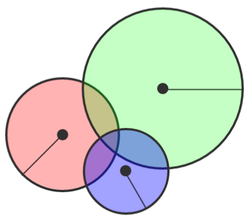
Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich

HLOC: Hints-Based Geolocation Leveraging Multiple Measurement Frameworks

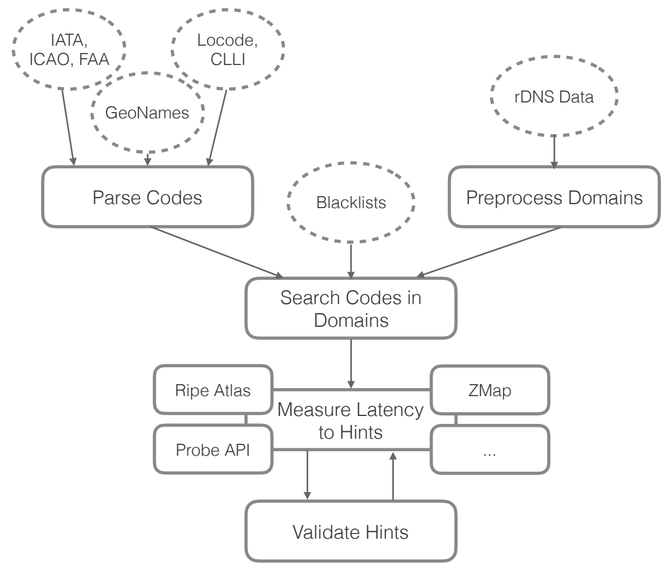
Patrick Sattler

Problem Statement



- Validate the location information in domain names
- Build a framework which:
  - is easy to use and scalable
  - combines different measurement technologies/frameworks
- Focus on router domain names
- Evaluate these with the chosen validation algorithm
- Compare the results with:
  - IP location databases (fast with unknown reliability [1])
  - Previous DNS-based approaches, e.g., DRoP [2]

Approach



- Modular architecture with independent steps

Validation Algorithm

$$dist(probe, host) < x \tag{1}$$

- **x** is the threshold distance between probe and suspected host location (we chose $x = 1000km$)
- From the set of the nearest probes a random one is selected

$$RTT(probe, host) < a + \frac{2 \cdot dist(probe, host)}{c \cdot c_0} \tag{2}$$

- **a** is the maximal buffer time (we chose $a = 9ms$)
- The analysis shows a linear growth of validated location hints

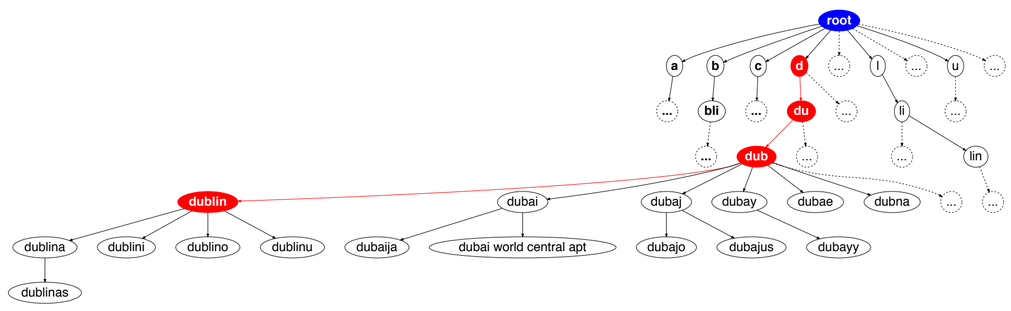
$$maxError(a, x) = 2 \cdot x + \frac{a \cdot (c \cdot c_0)}{2} \tag{3}$$

- Our maximal error is 2900 km
- 80% of the measurement probes used for validation are closer than 25km to the suspected location
  - For these measurements the maximal error is 950 km

Examples

- `be2590.rcr21.dub01.atlas.cogentco.com`
  - `rcr` Fulton City Airport in Rochester, Indiana US
  - `dub` Dublin, Ireland (1.9 ms)
- `te0-0-0-2.nr11.b020473-0.dub02.atlas.cogentco.com`
  - `dub` Dublin, Ireland (2.2 ms)
- `ae-0.facebook.amstnl02.nl.bb.gin.ntt.net`
  - ams (IATA): Amsterdam, Netherlands (2.3 ms)
  - face (ICAO): Ceres, South Africa
  - ace (IATA): Lanzarote, Spain
  - ceb (IATA): Lapu-Lapu City, Philippines
  - . . . (8 more matches)
- `cr-01.0v-00-04.anx32.nyc.us.anexia-it.com`
  - `nyc` (IATA): New York City, US (1.3 ms)
  - `anx` (IATA): Andenes, Norway

Trie Data Structure



- Contains all codes for a fast search process
- Returns all prefixes and all codes in the subtree of a key
- Labels are matched against the trie and all codes are found by slicing recursively the first character

Outlook

HLOC 2.0

- Improve the probe selection algorithm
- Measure more than one time per location per host
- Modularize the code architecture for better configuration
- Use a database for a flexible analysis and higher performance
- Integration into RIPE Atlas

Long Term goals

- Visualize possible location areas with the help of measurements
- Build a web service to geolocate host
- Integrate more measurement frameworks

Checkout github.com/tumi8/hloc for the source code
Contributions are welcome!

[1] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, "IP Geolocation Databases: Unreliable?," *ACM SIGCOMM Computer Communication Review*, 2011.
[2] B. Huffaker, M. Fomenkov, *et al.*, "DRoP: DNS-based Router Positioning," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 3, pp. 5–13, 2014.
[3] M. Zhang, Y. Ruan, V. S. Pai, and J. Rexford, "How DNS Misnaming Distorts Internet Topology Mapping.," in *USENIX Annual Technical Conference, General Track*, pp. 369–374, 2006.
[4] J. Chabarek and P. Barford, "What's in a Name? Decoding Router Interface Names," in *ACM Workshop on HotPlanet*, 2013.
[5] A. Edmundson, R. Ensafi, N. Feamster, and J. Rexford, "Characterizing and Avoiding Routing Detours Through Surveillance States," *arXiv*, 2016.
[6] Q. Scheitle, O. Gasser, P. Sattler, and G. Carle, "HLOC: Hints-Based Geolocation Leveraging Multiple Measurement Frameworks," in *TMA '17*, (Dublin, Ireland), 2017.

Patrick Sattler sattler@in.tum.de