IP Geolocation

TU Berlin
SS 2007
Florian Holzhauer

Overview

- Use cases
- Former approaches
  - Manual
  - Automatic
- Topology based geolocation
  - Idea
  - Improvements
- Comparison/ Evaluation

IP Geolocation

- Geolocation of an IP address
- Most current approaches by manually maintained database
  - Expensive
  - Inaccurate
  - Maintenance
- Triangulation techniques
  - Automatically
  - Inaccurate?

Online-Advertising

- Target audience
- Regionalized advertisements

Credit Card Fraud

- Suspicious pattern
  - Geolocation is only one pattern
- Usage of the credit card in different countries on same day
- Usage in non-typical country

Legal issues

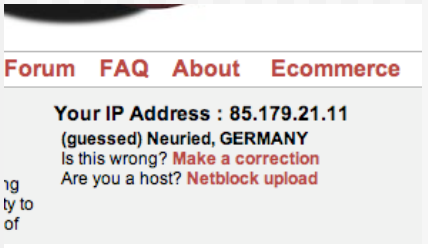
- US-Pharma-Ads are prohibited to advertise outside US
- TV/Audio with regional license
- Online-Casinos

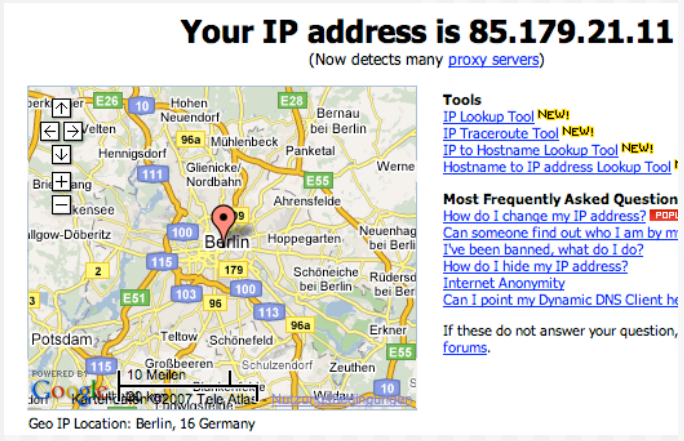
Emergency Calls with Voice over IP

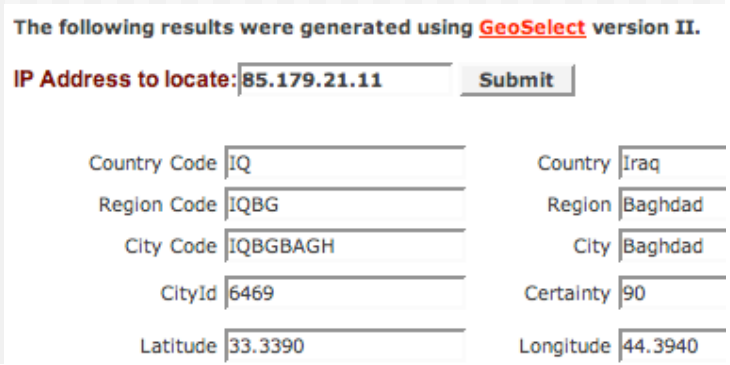
- „Röchelanruf“ - unconciousness
- High accuracy needed
- Most complicated use case

Manually maintained Databases

- Several commercial providers
- Community projects
- Problems
  - Outdated
  - Inaccurate
  - Not reliable
  - Expensive
- Benefits
  - Topology not important
  - Medium not important







Goals of automatic approaches

- Improve actuality
- Faster detection
- Cheap
- Usable in environments where manual work is not possible
  - Closed networks

Former automatic approaches

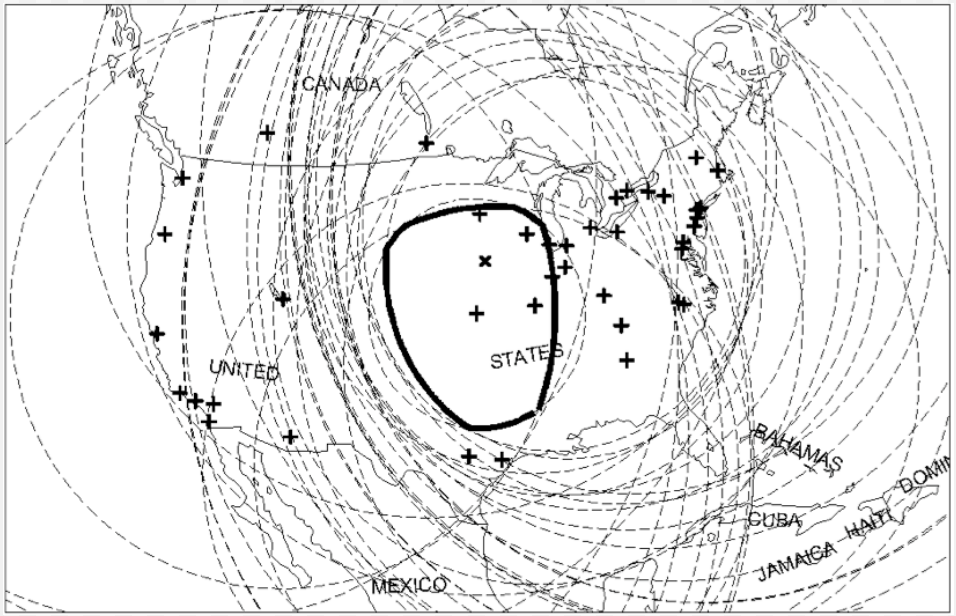
- Usually combination of several ideas
- „Triangulation“ via delay measurement
- Landmarks
  - Known location
- Targets
- PlanetLab
  - „PlanetLab currently consists of 805 nodes at 402 sites.“

GeoPing/Shortest Ping

- GeoPing: Two targets with similar latency are close to each other
  - Passive Landmarks
- Shortest Ping: Each target is assigned to the landmarks location with the smallest latency in between

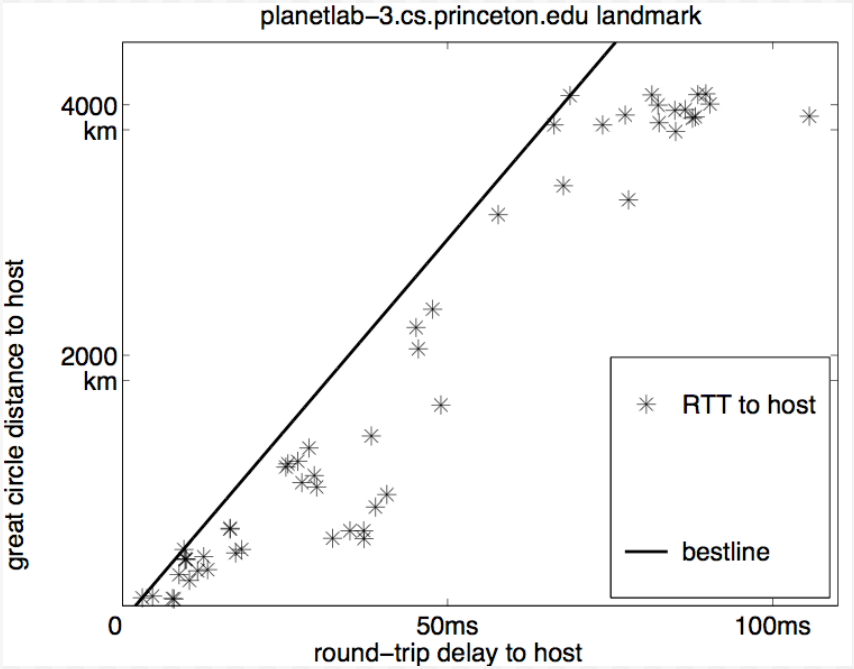
Constraint Based Geolocation (CBG)

- As GeoPing, but:
  - Initial measurement between all landmarks
  - Relation: Latency - physical distance
  - Triangulation of the target



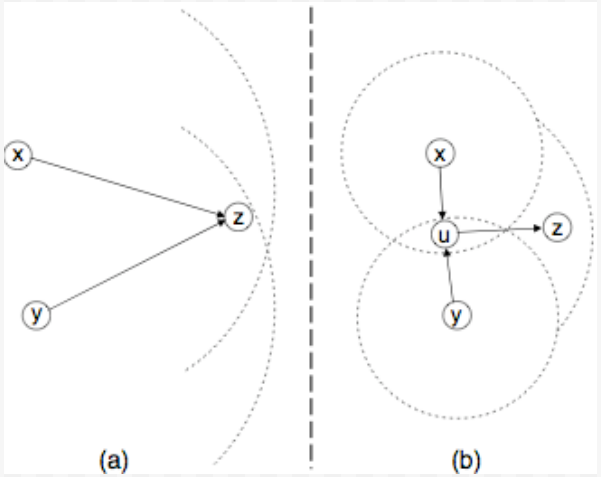
Speed of Internet (SOI)

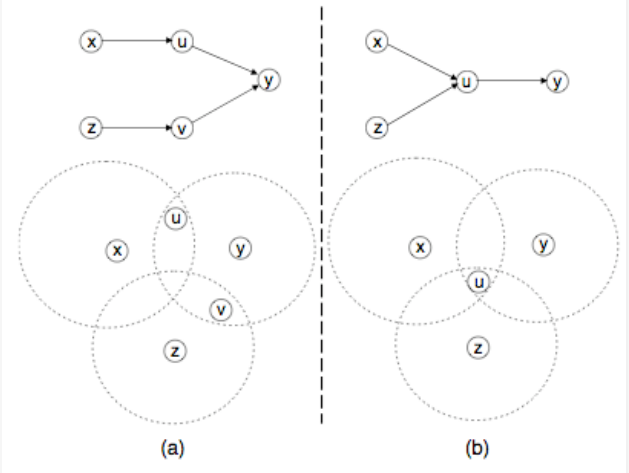
- Simplified CBG
- No initial measurements
- Fix „upper limit“ for the distance
- Theoretically 2/3 of speed of light in vacuum
- Practically: 4/9



Topology Based Geolocation (TBG)

- Improving resulty by topology knowledge
- Detection of aliased interfaces of a router





TBG steps

- Topology detection via traceroute
  - Both ways between landmarks
  - Estimate per-hop latency
  - Detect aliased interfaces
  - Location hints
  - Geolocate routers and targets simultaneously

Location Hints

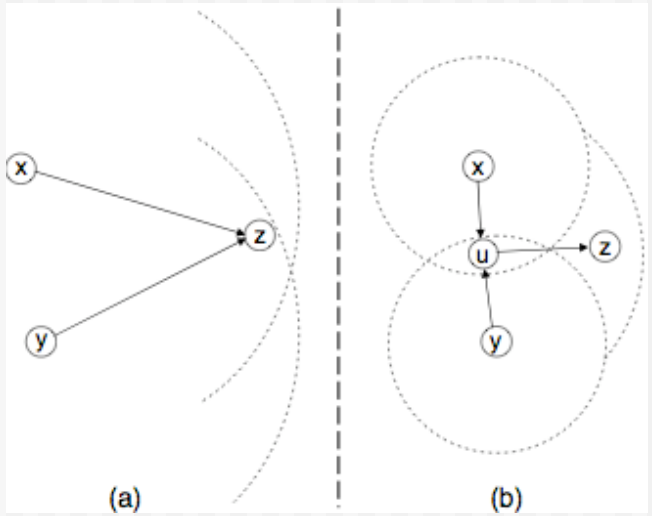
- Providerbased
  - Location „encoded“ in Reverse-DNS
    - Not a stand alone method
    - Used for validation
  - DNS Loc
    - Lat/Lon as DNS-Record
- Heuristic
  - High population density
    High probability

```
3  ge-4-1-0-101.cr02.ber.de.hansenet.
4  so-5-1-0-0.cr02.fra.de.hansenet.ne
5  po2-0.pr02.decix.de.hansenet.net
6  ffmxs11.decix.ffm.spxs.net (80.81.
7  gie5-1.ulmxs06.mu13.ulm.spxs.net
8  fe---1.r02.n25.obe.in-ulm.de (212.
```

```
;; ANSWER SECTION:
theseus.mathematik.uni-ulm.de. 86400 IN LOC     48 25 21.000 N 9 57 22.000 E 612.00
```

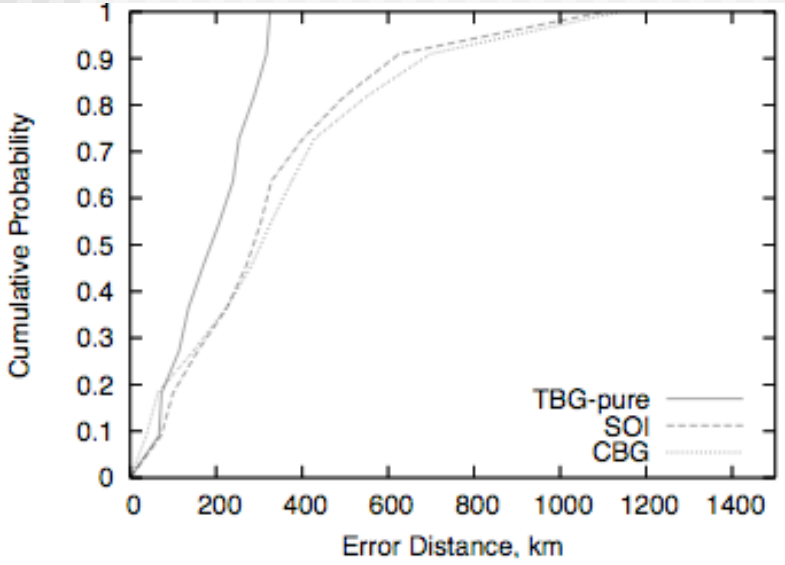
Hop Location

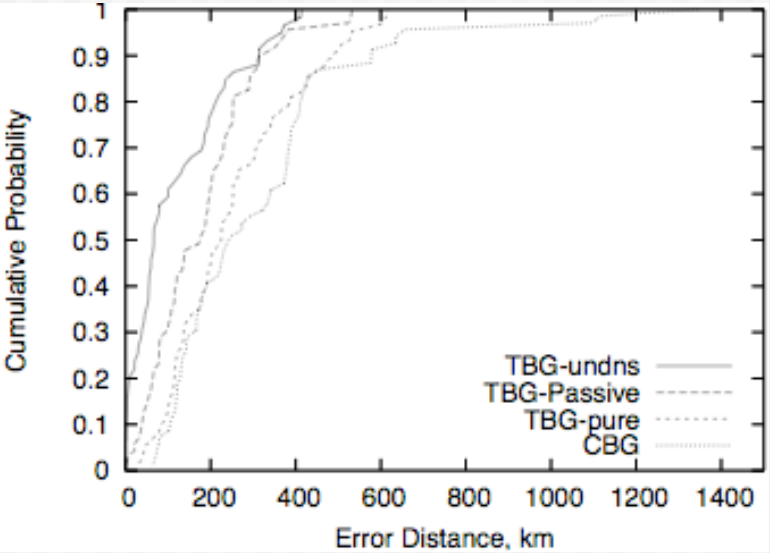
- Routers between landmark and target will be localised
  - Higher accuracy
  - Potential Passive Landmarks
- „Last Router“
  - Last multihomed router



Comparison

- TBG-pure: Only TBG
- TBG-Passive: + passive landmarks
- TBG-undns: Passive & Location Hints
- *SOI: Speed of Internet*
- *CBG: Constraint Based Geolocation*





Evaluation

- Three datasets
  - University network,, 2 Providers
- Median error
  - Test 1: CBG 689km, SOI 749km, TBG-pure 194km
  - Test 2: CBG 228km, TBG-pure 225km, -passive 176km, -undns 67km
- Location Hints:
  - 5509(*) of 8321 with parseable hints
- Alias Detection:
  - 2392 Alias-Pairs

(*) University Dataset, .edu domains nicht geparst

Summary

- Accuracy about 200km
  - Varies due to topology
- Accurate enough for most use cases
  - Useless for emergency calls
- Best results in combination of algorithms
- „Best algorithm“ varies for use cases
  - Initial measurements
  - Number of landmarks