Impact Factor (SJIF): 3.632

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

(Special Issue for ITECE 2016)

Different IP Based Geolocation Techniques to Determine the Location of the Host.

Mr. Kalpit Soni[1], Mr. Tejas Parmar[2], Mr. Bhavin Patel[3]

1. *Ph.D. Research Scholar, Charotar University of Science and Technology, Changa, Gujarat, India*
2. *Assistant Professor, Shri A.N.Patel P.G.Institute, Anand, Gujarat, India*
3. *Assistant Professor, Shri A.N.Patel P.G.Institute, Anand, Gujarat, India*

Abstract: Many applications benefit from using IP geo-location to determine the geographic location of hosts on the Internet. In cloud computing, people get concerned not only of privacy and integrity of their data but its location as well. This is important because several of laws and regulations governing data storage and access mechanisms in different geographical regions. In cloud applications some organizations must ensure their virtual machines stay in an appropriate geographic region. This paper studies basic of geolocation, different IP geolocation based techniques with their detailed methodology of working, pros and cons of existing techniques of IP geolocaiton. This research paper covers Shortest Ping, GeoPing, CBG, Statistical, Learning Based, TBG and Octant. All these methods are used for finding the current geographic location of the Internet hosts.

I. INTRODUCTION

Many applications benefit from using IP Geo-location to determine the geographic location of hosts on the Internet. For example, Online content providers and search engines tailor their content based on the client's location. Geolocation also used in many security sensitive applications. Online Content provider distribute to specific geographic regions. Before allowing a clinet to view the content, they determine the client's location from its IP address and allow access only if the client is in a permitted jurisdiction.

The growth of infrastructure as a service clouds, such as Amazon's Elastic Cloud Computing, there are many organizations that could benefit from cloud computing. Users of cloud computing deploy VMs on a cloud provider's infrastructure without having to maintain the hardware their VM is running on. However, differences in laws governing issues such as privacy, in- formation discovery, compliance and audit require that some cloud users to restrict VM locations to certain jurisdictions or countries. These location restrictions may be specified as part of a Service Level Agreement between the cloud user and provider. Cloud users can use IP geolocation to independently verify that the location restrictions in their cloud SLAs are met.

IP geolocation has been an active research for almost a last decade. However, all current geolocation techniques assume a benign target that is not trying to intentionally mislead the cloud user, and there has been limited work on geolocating malicious targets.

II. GEOLOCATION BACKGROUND

IP geolocation aims to solve the problem of determining the geographic location of a given IP address. The solution can be expressed to varying degrees of granularity; for most applications the result should be precise enough to determine the city in which the IP is located, either returning a city name or the longitude and latitude where the target is located. The two main approaches to geolocation use either active network measurements to determine the location of the host or databases of IP to location mappings.

III. GEOLOCATION TECHNIQUES USING DATABASES OF IP.

The most popular approach to identify the location of an IP is to keep and manually maintain a databases. The databases are used to match an incoming IP address to the country, region, state, city, latitude, longitude and Internet Service Provider of the internet user. These databases can be either proprietary or public. Public databases include those administered by regional Internet registries. (ARIN, RIPE). Proprietary databases of IP to geographic location mappings are provided by companies such as MaxMind[1], IP2Location[2] or Neustar IP Intelligence.[3] Registries and databases tend to be coarse gained, usually returning the headquarters location of the organization that registered the IP address. This

---

[1] http://www.maxmind.com.

[2] http://www.ip2country.net/

[3] https://www.neustar.biz/services/ip-intelligence

becomes a problem when organizations distribute their IP address over a wide geographic region, such as large Content Providers. Human errors should also be taken into consideration since the risk or entering accidently or intentionally wrong data is increased.

IV. GEOLOCATION TECHNIQUES USING MEASUREMENT BASED

An alternative to databases of IP geolocation is Measurement-based geolocation algorithms. Measurement-based geolocation is particularly appealing for secure geo-location because if a measurement can reach the target even if it is behind a proxy, the effectiveness of proxying will be diminished. Based on reported accuracies, the current geolocation algorithms are sufficiently accurate to place a host within a country or jurisdiction. Much research has gone into improving the accuracy of measurement-based geolocation algorithms.

Class	Algorithm	Average Accuracy (KM)
Delay-based	GeoPing	150
	CBG	78-182
	Statistical	92
	Leaning-based	407-449
Topology-aware	TBG	194
	Octant	35-40
Other	GeoTrack	156

Table 1: Average accuracy of measurement-based geolocation algorithms.

Geolocation algorithms mainly rely on ping and traceroute measurements. Ping measures the round trip time (RTT) delay between two hosts on the Internet, while traceroute discovers and measures the RTT to routers along the path to a given destination. We can classify measurement-based geolocation algorithms by the type of measurements they use to determine the target's location.

A. Shortest Ping

Shortest Ping was one of the earliest attempts to use Internet measurements to geolocate a target host. In this method, a set of hosts, called landmarks, perform network delay measurements by transmitting ICMP ping packets between each other. When a new target is encountered, the landmarks determine their delays to the target. These delays are compared with the existing measurements. In particular, let L denote the index set for Landmarks, i.e., the set of landmarks is given by {Li : i ∈ L}. The location of each Landmark Li, where i ∈ L, denoted by (ϕi , λi), is assumed to be known. Here, ϕi and λi represents the longitude and latitude, respectively, of Landmark Li in units of radians. Let d_{IT} denote the delay from Landmark i to the target T. In Shortest Ping, the location estimate for the target is defined as (φk, λk), where

$$k = \arg\min_{i \in L} \{d_{IT}\}$$

Shortest Ping depends only on the minimum RTT delay, an inaccurate measurement or a high speed link may have a significant impact on the estimated target location.

B. Geoping

GeoPing tries to convert network delay to geographic constraints. Given landmark measurement point to destination network delay, assume the speed of signal transported in line is the maximum of the light. The limit distance between target and the destination is the radius to a landmark position as the center of circle. With more landmarks, the public area of the circles is the target place. But in fact, due to propagation speed limit was set at 2/3 the speed of light by researchers which lead to the radius of circle larger. The overlapping circular portion is large too, so GeoPing can't locate it accurately.

C. Constraint Based Geolocation

To describe the Constraint Based Geolocation technique, we need to define two terms:

1. Multilateration is a method where the position of a point can be estimated using a sufficient number of distances to some fixed points, whose positions are known.
2. Additive distance distortion: due to imperfect measurements, delay is distorted additively with respect to the time for light to pass through the path.

The main idea in CBG is to estimate the distance between the Landmarks and the target using delay measurements, create constraints based in these distances, and finally apply them in a multilateration process to geolocate the Target. In order to understand the intuition of this technique, we will use an example taken from[4][5][6]. Let's say that we have a set of landmarks L = {L_1, L_2, L_3}, and a target T. Each Landmark applies its additive distance constraints to the target G_{iT} , which is given by the addition of the additive distance distortion ($\gamma i\tau$) and the real geographic distance (G_{iT}): $\hat{G}_{iT} = G_{iT} + Y_{iT}$. he estimated location of T will be in the intersection of those distances Figure1.

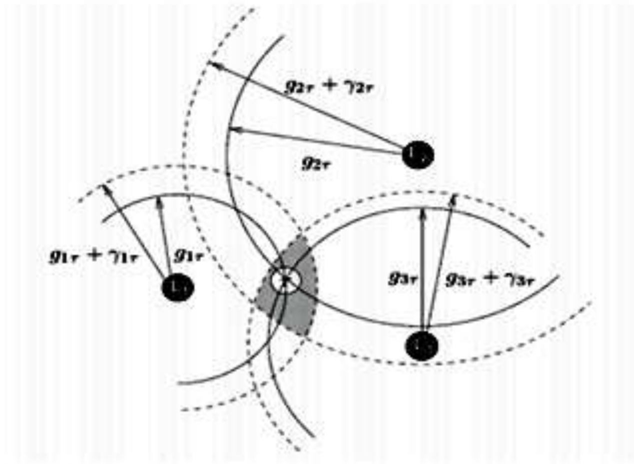

Figure 1. Multilateration with geographic distance constraints.

Interpretation of delay measurements to distance constraints: The last thing that has to be addressed here is how they transformed delay measurements into the distances. In Figure2 shows the calibration between delay over distance for 3 given Landmarks. The additive distance constraints can be can be extracted by calculating the best line for each landmark. When those distances are known, using this calibration the location of the target can be determined.

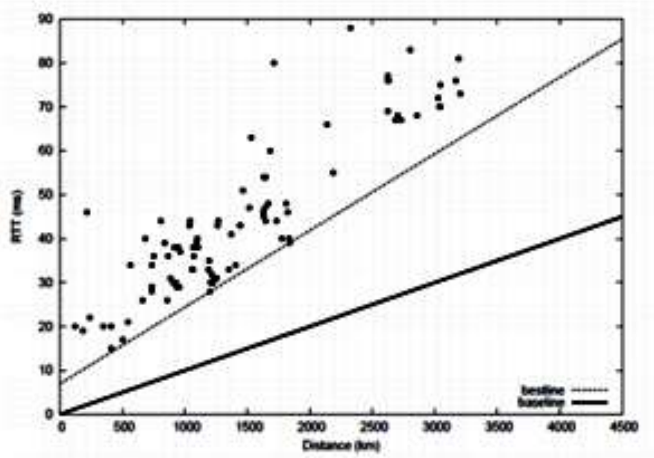

Figure 2. Sample scatter plot of geographic distance and network delay.

In the geolocation for the target, the TBG methods also leverage the network topology in addition to the relationship between Internet delay and geographic distance.E Katz-Bassett et al[6]. have taken into account the network topology. Owing to the utilization of the relationship between Internet delay and geographic distance, both of the delay-based methods and topology based methods may not really work in two conditions: (1) the delay from the target to the landmark deviates much from the normal value; (2) one malicious Internet host, known as adversary, tries to cover his geographic location by tempering with the Internet delay measurements. In that case, the geographic location of the Internet host appears to be incorrect, which is presented as an instance in Figure 2.

[4] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based geolocation of internet hosts. IEEE/ACM Trans. Netw., 14(6):1219– 1232, 2006.

[5] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based geolocation of internet hosts. In IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pages 288–293, New York, NY, USA, 2004. ACM.

[6] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based geolocation of internet hosts. IEEE/ACM Trans. Netw., 14(6):1219– 1232, 2006.

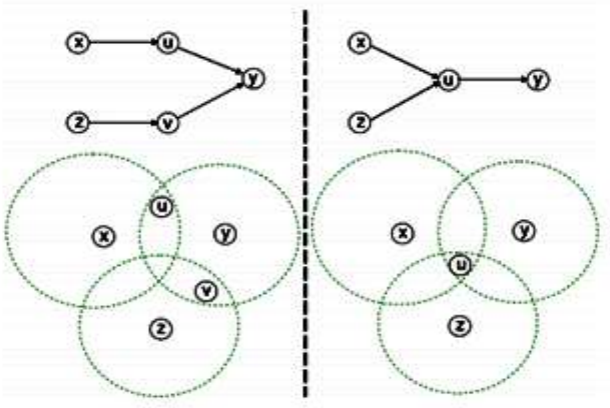

Figure 3. Router aliases: Accuracy improves from (A) to (B) as v is identified as an alias for u.

- Use measurements to extract existing structural constraints
- Use delay and topology measurements to validate the locations of passive landmarks and external location hints and then incorporate them to overcome insufficient structural constraints.

The scheme: In the beginning the Landmarks probe traceroutes to each other Landmarks and the target. This gives both the round trip time and the identity of the intermediate hops. An estimation of link latencies is provided through the differences between adjacent network interfaces, while collocated interfaces are also identified. Finally, combining the entire set of traceroutes with structural observations provides us the network topology. Since we know the topology, the target is geolocated along with all of the intermediated routers using a constrained based optimization technique. At the end, there will be multiple constraints on the positions of the intermediate routers forcing them to be placed close to their actual positions. These placements provides the geographical detours taken by end to end paths and resulting in a good estimation of the targets position. finally, TBG:

- Takes advantage of the fact that routers nearby landmarks are easy to locate.
- Uses the locations of intermediate routers to quantify the directness of paths to targets, thus making these measurements more useful.
- Allows the solution to be iterated using the coupling between network elements until all the elements have converged to an overall map that is consistent, as it must be in reality.

D. Octant

This approach transforms the problem to an error minimizing constraint satisfaction. The constraints are derived from network measurements. Octant imposes even more constraints than CBG. The system of constraints is solved using geometric methods. The constraints $\gamma 0,...,\gamma n$ can be either positive or negative. Each constraint defines a region where the target node is believed to be located. This space is represented geometrically as a region βi bounded by Bezier curves where node i is estimated to be located as shown in Figure 4. "Bezier curves are polynomial parametric curves with n + 1 control points P_0, . . . , Pn where n in the order of the polynomial with n=3 for most implementations. Intuitively, the start point P_0, end point Pn with the remaining points providing the directional information ."[7]

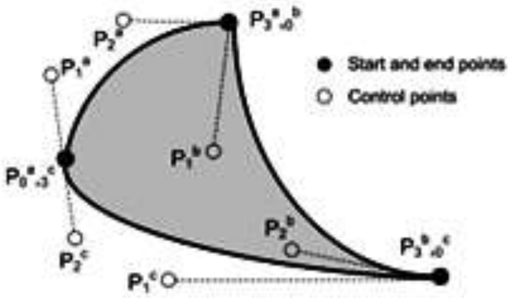

Figure 4. "Location representation in octant".

---

[7] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards ip geolocation using delay and topology measurements. In IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, pages 71–84, New York, NY, USA, 2006. ACM

Octant represents the expected target location as a region bounded by a set of Bezier curves. Each curve a, b, c consists of four control points P0, . . . , P3 with P0 and P3 as the start and end points respectively and P1 and P2 as control points that help direct the curve. Bezier curves provide a compact way to represent large, complex areas precisely. They also admit efficient union, intersection, and subtraction operations.[8]

The constraints are obtained using a number of Landmarks whose positions are known. Given a set of Ω of positive constraints and Φ is a set of negative constraints that define the location of a target node T the estimated location region for the target is given by:

$$\beta_T = \bigcap_{X_T \in \Omega} X_T \setminus \bigcup_{X_T \in \Phi} X_T$$

Octant determines the correlation between the geographical distance of any pair of Landmarks and the corresponding network measurements. The functions that correspond to the upper and lower bounds of the convex hull are computed. Particular these functions tight and conservative constraints can be extracted for the latency-to-distance mappings. The framework outputs a set of points where the target may be located and this framework can integrate additional constraints from the previous approaches such as locations of IPs stored in a database or locations extracted from GeoTrack in to order to improve its performance.. Even though in all the experiments that were conducted in show that the accuracy of this method is very high it is important to mention that it did not perform well where Octant was used in a Metropolitan area network.

V. CONCLUSION

In this paper we have discussed the existing techniques are used to estimate the geographic location of internet hosts. Geolocation of IP can be achieved by means of internet measurements or by using third-party based utility. Depending on the accuracy that is expected by the application a different techniques are used. The measurement based and topology aware techniques are sophisticated methods because these are independent from the third-party based database. Granularity of CBG and Octant is up to city level, but Octant sometimes not performs well in Metropolitan Area Network.

[8] B. Wong, I. Stoyanov, and E. G. Sirer. Geolocalization on the internet through constraint satisfaction. In Proceedings of Workshop on Real, Large Distributed Systems (WORLDS). Workshop on Real, Large Distributed Systems (WORLDS), November 2006.