Survey of Different IP Based Geo Location Techniques

Darshan Patel[1] Mrs. Gargi Chauhan[2]
[1]M.E. Student [2]Assistant Professor
[1,2]Department of Information & Technology
[1,2]SVIT, Gujarat, India

Abstract—Many applications benefit from using IP Geolocation to determine the geographic location of hosts on the Internet. This is a challenging problem because the IP addresses do not structurally contain information about the corresponding geographic locality. By knowing the location of consumers in real time, web sites can display localized content, server bandwidth balancing, improve click-through and sales, prevent Internet frauds, and implement many other solutions. This survey paper shows the study over basics of geolocation, different IP geolocation based techniques with their detailed methodology of working, pros and cons of existing techniques of IP geolocation. The methods, which we have reviewed in this paper are Shortest Ping, GeoPing, CBG, Statistical, Learning based, TBG and Octant. These methods are used for finding out the geographic location of the internet hosts.

Key words: Geolication, Multilateration, CBG, Octant, TBG

I. INTRODUCTION

Many applications take advantage from using IP Geolocation to estimate the geographic location of hosts on the Internet. Geolocation is also used in security-sensitive applications. For example, and search engines tailor their content based on the users's location. Online content providers like ReadMedia, BBC iPlayer, Hulu and Pandora limit their content distribution to specific geographic regions. Before allowing a client 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.

Looking ahead, the growth of IaaS clouds, such as Amazon's EC2 service, may also run organizations using cloud computing to user's geolocation. Users of cloud computing deploy VMs on a cloud provider's infrastructure without having to maintain the hardware their VM running on. However, it is required to restrict some cloud users by accessing VM locations to certain jurisdictions or countries regarding difference in law governing issues such as privacy, information discovery, compliance and audit. These location restrictions may be specified as part of a service level agreement (SLA) between the cloud user and provider. Cloud user can use IP geolocation to independently verify that the collation restrictions in their cloud SLAs are met [12].

In these cases, the target of geolocation has an incentive to mislead the geolocation system about its true location. Clients commonly use proxies to mislead content providers so they can view content that is unauthorized in their geographic region. In response, some content providers however, have identified and blocked access from known proxies; but this does not prevent all clients from circumventing geographic controls. Similarly, cloud providers may attempt to violate location restrictions in their SLAs to move customer VMs to cheaper locations. Governments that enforce location requirements on the cloud user may require the geolocation checks to be robust no matter what a cloud provider may do to mislead them. Even if the cloud provider itself is not malicious, its employees may also try to relocate VMs to locations where they can be attacked by other malicious VMs. Thus, while cloud users might trust the cloud service provider, they may still be required to cd. have independent verification of the location of their VMs to meet audit requirements or to avoid legal liability[12]. There are two main approaches to geolocation

1) Databases of IP
2) Active Network Measurements to determine the location of the hosts.

II. GEO LOCATION TECHNIQUES USING DATABASES OF IP

Keeping and manually maintain a databases is the most popular and basic approach to identify the IP location. 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 [3], IP2Location [4]. Registries and databases tend to be coarse gained, usually returning the headquarters location of the organization that registered the IP address. This 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.

III. GEO LOCATION TECHNIQUES USING MEASUREMENTS BASED

Measurement-based geolocation algorithms are an alternative to databases of IP geolocation. Measurement-based geolocation is particularly appealing for secure geolocation 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 machine within a country or jurisdiction. Many researches have gone into improving the accuracy of measurement-based geolocation algorithms up city level or street level.

A. Shortest Ping

Shortest Ping (SPing) was one of the earliest attempts to use Internet measurements to geolocating a target host. In SPing, a set of hosts, called landmarks, perform network delay measurements by transmitting ICMP ping packets between each other. When a new target host is encountered, the landmarks determine their delays to the target. These delays are compared to the existing measurements.

Class	Algorithm	Average Accuracy(km)
Delay Based	GeoPing	150 km (25th percentile); 109 km (median)
	CBG	78-182
	Statistical	92
	Learning Based	407-449
Topology-Aware	TBG	194
	Octant	35-40
Other	GeoTrack	156

Table 1. Average accuracy of measurement-based geo location algorithms

More precisely, let L denote the index set for landmarks, i.e., the set of landmarks is given by $\{L_i : i \in L\}$. The location of each landmark L_i , $i \in L$, denoted by (ϕ_i , λ_i), is assumed to be known. Here, ϕ_i and λ_i represents the longitude and latitude, respectively, of landmark L_i in units of radians. Let $d_{i\tau}$ denote the delay from landmark i to the target τ . In Shortest Ping, the location estimate for the target is defined as (φ_k, λ_k), where

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

Since SPing 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 two thirds 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 Geo Location*

In order to describe the CBG technique, we need to define two terms:

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.

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 targets using delay measurements, create constraints based in these distances, and finally apply them in a multilateration process to geolocate the targets. In order to understand the intuition of this technique, we will use an example taken from [7], [6]. Lets say that we have a set of landmarks L = {L1, L2, L3}, and a target τ . Each landmark applies its additive distance constraints to the target $g_{i\tau}$, which is given by the addition of the additive distance distortion ($\gamma_{i\tau}$) and the real geographic distance ($g_{i\tau}$): $\hat{g}_{i\tau} = g_{i\tau} + \gamma_{i\tau}$. he estimated location of τ will be in the intersection of those distances Figure1.

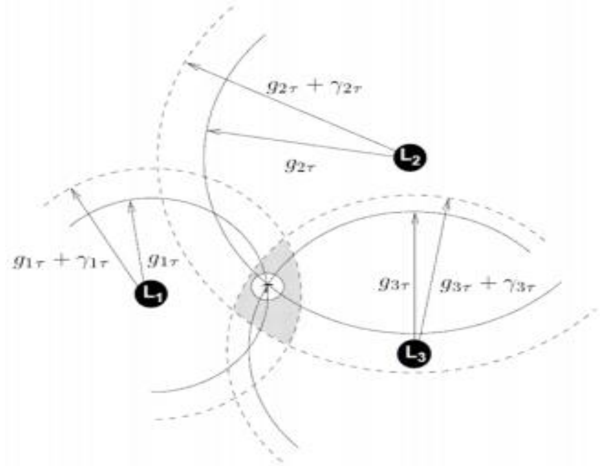

Fig. 1: Multilateration with geographic distance constraints. [7]

Interpretation of delay measurements to distance constraints: The last thing that has to be addressed here is how they transformed delay measurements into distances. In Figure2, there is a plot of delay times over distance for a 3 given landmark. The additive distance constraints can be can be extracted by calculating the best line for each landmark. When those distances are known, the constraints are applied and the location of the target can be determined.

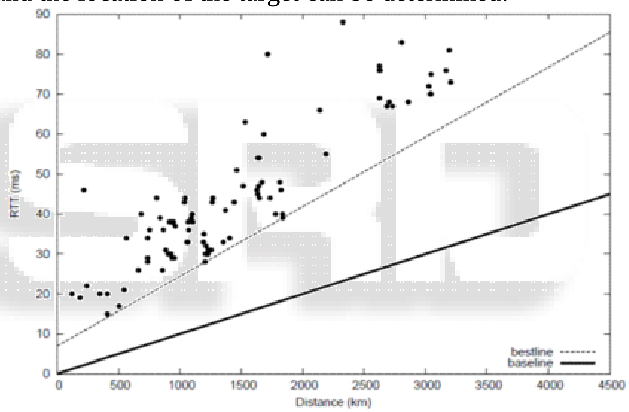

Fig. 2: Sample scatter plot of geographic distance and network delay. [7]

D. *TBG*

In this technique, in order to calculate the geographic location of an IP, E Katz-Bassett et al. [8] have taken into account the network topology. It is an extension of CBG in 4 a way that (a) it augments delay measurements to end-hosts with information about topology and routing and (b) it performs constraints multilateration in all the intermediate routers and end-hosts (targets) simultaneously. The proposed scheme is motivated by the following observations that a geolocation technique must satisfy: • Take into account network topology and routes in order to capture path-specific latency inflation's.

Simultaneously geolocate the targets as well as routers encountered on paths from landmarks to other landmarks or targets.

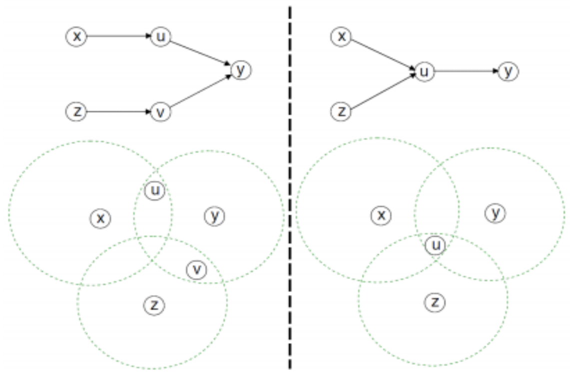

Fig. 3: Router aliases: Accuracy improves from (a) to (b) as v is identied as an alias for u. [8]

Use measurements to extract existing structural constraints (collocated interfaces Figure5 [8])

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 and the target. This gives both the RTT 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-toend paths, resulting in a good estimation of the targets position. Conclusively, 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.

E. 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, in other words these constraints define where the node can and cannot be. Each constraint defines a region where the target node is believed to be located. The solution space is represented geometrically as a region βi bounded by Bezier curves where node i is estimated to be located as shown in Figure4. "Bezier curves are polynomial parametric curves with n + 1 control points P0,..., Pn where n in the order of the polynomial with n=3 for most implementations. Intuitively, the points P0 and Pn are the start and end points with the remaining points providing the directional information [9]."

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

$$\beta_\tau = \bigcap_{X_\tau \in \Omega} X_\tau \setminus \bigcup_{X_\tau \in \Phi} X_\tau$$

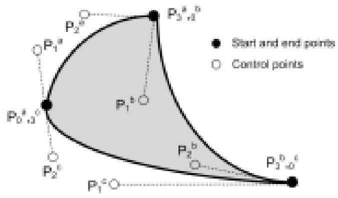

Fig. 4: "Location representation in Octant

Octant represents the estimated 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. This gure requires a total of only nine control points to precisely dene. Bezier curves provide a compact way to represent large, complex areas precisely. They also admit efficient intersection, union, and subtraction operations." [9]

A positive constraint represents that a node is within d miles of a landmark L and geometrically can be represented as a disk with radius d centered at the landmark. A negative constraint restricts that the target is not located within distance d which is the complement of a positive constraint. The area may be non-convex and may consist of disconnected regions. Octant uses a weighted solution technique where weights correspond to confidence in the corresponding constraint. Since Octant use aggressive along with conservative constraints the constraint system always has a solution.

Octant periodically determines the correlation between the geographical distance of any pair of landmarks and the corresponding network measurements. It determines the convex hull around the points on the graph. The functions that correspond to the upper and lower bounds of the convex hull are computed. Given 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. 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. Geographical constraints such as oceans and deserts can be expressed in Bezier-regions constraints and added in the framework

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 in [5] where Octant was used in a Metropolitan area network.

IV. 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.

REFERENCES

[1] http://sahara.cs.berkeley.edu/rawtraces/.
[2] http://www.hostip.info/
[3] http://www.ip2country.net/
[4] http://www.maxmind.com.
[5] R. Baden. Ip geolocation in metropolitan area networks.
[6] 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.
[7] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based geolocation of internet hosts. IEEE/ACM Trans. Netw., 14(6):1219– 1232, 2006.
[8] 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
[9] B. Wong, I. Stoyanov, and E. G. Sirer. Octant: A comprehensive framework for the geolocalization of internet hosts. In NSDI, 2007
[10] 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.
[11] R. Baden. Ip geolocation in metropolitan area networks GILL, P., GANJALI, Y., WONG, B., AND LIE, D. Dude, where's that IP? Circumventing measurement-based IP geolocation. In Proceedings of the USENIX Security Symposium (2010).
[12] Doxa Chatzopoulou and Marios Kokkodis. IP Geolocation,https://www.researchgate.net/publication/228453630_IP_geolocation.

