Revisiting Constraint Based Geo Location: Improving Accuracy through Removal of Outliers

Sameer Qazi and Muhammad Kadri
College of Engineering, Karachi Institute of Economics and Technology, Pakistan

Abstract: *IP based GeoLocation has become increasingly important for network administrators for a number of reasons. The first and foremost is to determine cyber attackers' geographical location to prevent, contain and thwart further attacks. A secondary reason may be to target potential advertisements to users based on their geographical location. IP addresses are indicators of the users location at a course level such as country or city. For further refined estimate of user location within a city, for example, network measurements as simple as simple round trip time of pings to the user can give a more precise estimate of users location. Recently, several researchers have proposed constraint based optimization approaches to estimate a users location using a set of landmark hosts using multi-lateration. In this work, we show that the optimization of such schemes are sometimes plagued by outliers which causes location estimated to be deteriorated greatly. We provide anecdotal evidence for this and proposals to alleviate these problems for detection, masking or removal of these outlier measurements and show that location error improvement of several hundreds or thousands of km are possible.*

Keywords: *Geo location of internet hosts, constraint-based optimization, location error.*

Received November 17, 2014; accepted March 23, 2015

1. Introduction

With the recent increase in threat of high profile cyber crime activities against banks, military, stock markets, corporate organizations and metropolitan Supervisory Control and Data Acquisition (SCADA) systems; it is of immense importance to quickly pinpoint the geographical location of attacker operating from otherwise seemingly benign internet host. The location of the attacker can be revealed if the computer terminal he is operating from can be pinpointed. Inspired by the current research in position tracking of mobile hosts; such as outdoor location tracking using Global Positioning System (GPS) [6] using trilateration of several satellite signals to find the geographical co-ordinates of the target and indoor location tracking using WiFi based Positioning Systems (WFPS) using Received Signal Strength (RSS) [16] from one or several wireless Access Points (APs). Recently, several researchers have proposed similar techniques for pinpointing the geographical location of Internet Hosts using multilateration from other trusted Internet hosts. Since multilateration in such a manner would result in very crude estimates in absence of sophisticated electronics like GPS and knowing the facts that signals on the Internet do not take the shortest path; they are restricted to travel through pre deployed fiber optic /copper cables and are also governed by routing policies of Internet Service Providers (ISPs). These artifacts are catered for by solving a constraint based optimization problem to estimate a cyber attackers' location after obtaining multilateration results from multiple vantage points in the network.

Constraints generalize the rules on how signals can and cannot propagate keeping in view both the laws of physics and man-made restrictions. Satisfying constraints arbitrarily might result in a large solution space, which can be tightened by using an objective function that is to be optimized (minimized) while satisfying all constraints.

In this work, we show that such schemes are sometimes plagued by outlier which causes location estimates to be deteriorated greatly. We provide anecdotal evidence for this and proposals to alleviate these problems for detection, masking or removal of these outlier measurements and show that location error improvement of several hundreds or thousands of km is possible.

The rest of this paper is as follows: section 2 describes related work, section 3 discusses the shortcomings and challenges in using constraint based optimization schemes for geo-locating Internet hosts. Section 4 discusses schemes for outlier measurements detection and their masking or removal; section 5 discusses the results of the proposals and contributions of the paper. Section 6 finally concludes the paper.

2. Related Work

Gueye *et al*. [9] were the first research group to investigate the problem of geolocating Internet hosts using Constraint Based Geolocation (CBG) approach which is a solution to constraint based optimization problem. They showed that optimization problems satisfying the basic laws of the physics of signal propagation speeds and queueing delays at routers can indeed help predict the geographic location of Internet hosts, albeit with some prediction errors. Internet based co-ordinate schemes involving only round trip time measurements between a set of landmark host forming a *Lp* measurement space has been previously discussed by several researchers.

More recently Dong *et al*. [8] have approached the problem of IP based geolocation by doing a real-time prediction of host distance; the approach involves clustering to divide data in different bands and application of polynomial regression model for more accurate estimation of queueing delays.

Geolocating Internet hosts may rely heavily on network measurements which may not be scalable [1] as the overheads scale as $O(N^2)$ for pairwise measurements between N landmark hosts, therefore several researchers [3, 4, 19, 20] turned their attention towards optimizing network probing overheads by proposing statistical prediction models when the topology of the network is known in the ideal case as well as when topology of the network is only known partially or even incorrectly.

Wang *et al*. [23] have similarly identified a new methodology where network measurements potentially give crude estimates of geographic location with errors in the ranges of hundreds of kilometers; these can be drastically reduced to within half a mile when it is coupled with location information often advertised by businesses on their locally stored websites. Similar approach has been adopted by Katz-Bassett *et al*. [13] in an scheme outlined as Topology Based Geolocation (TBG) where topology based information is also exploited for improved accuracy. Wong *et al*. [24] have developed a framework “Octant”, where both positive and negative constraints are used to accurately determine the position of the host but it again exploits some topology information to fine tune the location of the host. Our work is different from these, as we solely try to infer location estimates without using any topology information which is often misleading and hard to get. Routers are known to drop traceroute requests; Traceroutes are known to suffer from many artifacts as described by Augustin *et al*. [2] recently in their Paris Traceroute architecture. Topology information can often undergo huge changes in case of inter domain routing events as investigated recently by the PathMiner architecture by Comarela and Crovella. [5]

Other researchers have addressed the problem of cleaning the data by removal of outlier measurements for better analysis of conforming data [14, 15]. They have devised statistical methodologies to detect, and remove outlier measurements for better analysis of remaining data. Our work follows this approach of outlier removal.

3. Challenges in Constraint Based GeoLocation

Gueye *et al*. [9] presented their ground breaking work in CBG of host using just Round Trip Time measurement and Latitude-Longitudes between a set of landmark hosts amongst themselves which could be used to predict the geographic location of an unknown host (Figure 1). The idea is based on that fact that round trip time estimate between any two internet host primarily involves the time spent in travelling between the two hosts, a theoretical speed of 1ms round trip time per 100km distance. The only other factors contributing on any extra time is the queueing and transmission delays at router input/output queues and packet lookup at routers. If we can specify all of these as a fixed time constant, the propagation model of the signal over the round trip between any two internet hosts is just given by the simple linear relationship of:

$$\begin{aligned}&\text{Round Trip Distance Estimate} = \\ &\text{Signal Propagation Speed} * (\text{measured rtt} - \text{router delays})\end{aligned} \tag{1}$$

The above problem formulation assumes that:

a. Internet uses the shortest straight line geographical path between any two computer hosts. This is necessarily not true as signals travel though shared long-haul cables between continents and also must pass through intermediate waypoints at the router locations to determine next hop. Hence, a circuitous path may be followed.
b. Routers take some constant average time to process the packets which is mainly reading the destination IP address field in packet headers and consultation of their routing tables stored in memory for the best possible next hop through partial match classification. Technically, router delays are not fixed. The router delays vary and the primary reason is that network traffic does not stay constant, it varies with time of day effects as well as flash crowds. The Internet traffic model and the queueing delays associated with it at routers is widely accepted as behaving according to the M/M/1 queueing model, where the packet arrivals at routers are governed by poisson process or exponential inter-arrival times and the processing times (lookup and mainly transmission delays) are also distributed according to exponential distribution due to such distributions of packet sizes. Dong *et al*. [8] addressed this issue by developing a framework for real time internet host location estimation.
c. The round trip delay measurements to the unknown host from the landmarks are accurate; this is only possible if the unknown host is not using some disguising technique to hide its true IP address. If the unknown host is using some approach to hide its true IP address then in order to determine its

location some preliminary technique would first have to be used to determine its true IP address to obtain accurate estimates of round trip times from the landmark hosts to the unknown host. Such techniques are outside the scope of this paper, which focusses more on alleviating the effects of outlier measurements on the original CBG technique.

Since in the above relationship what is generally known are a set of readings amongst some landmark nodes, namely the distance and round trip times. Distance is estimated from the latitude and longitude positions of landmark hosts using the earth geometry while round trip times are actually measured and averaged (sections 4 and 5). These readings are then used to calculate for each landmark a reasonable linear relation between distance and time characteristics according to the equation shown above. An unknown host's location may be estimated by only a round trip time measurement to it by all the landmark hosts determining circular regions of Interest for each. Finally the location of an unknown host can be established through multilateration; i.e., a common overlapping area of Regions of Interest of all the landmarks. The centroid of the resultant polygon with vertices at the intersection points of the common Region of Interest is an estimate of the true geographic location of the unknown host (Figure 1).

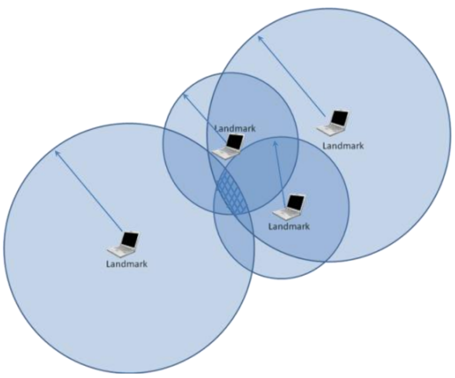


Figure 1. Using multilateration to estimate the position of an unknown host showing by the overlapping region of regions of interest of all the landmarks.

Several approaches are possible to process the distance-time relationships of each of the landmarks to estimate the circles in which the unknown host should lie. The simplest one would be to use a linear regression model to estimate the distance time relationship for each landmark and then project it to find the location of an unknown. However, a regression line will roughly only satisfy the distance/ time relationship of half of the points of the landmark; either the points in the region above the line or below it. So it cannot be expected to geolocate the unknown host with 100 percent chance.

Instead, Gueye *et al*. [9] have proposed that a constraint based optimization problem CBG be solved to estimate the best line for each landmark Figure 2.

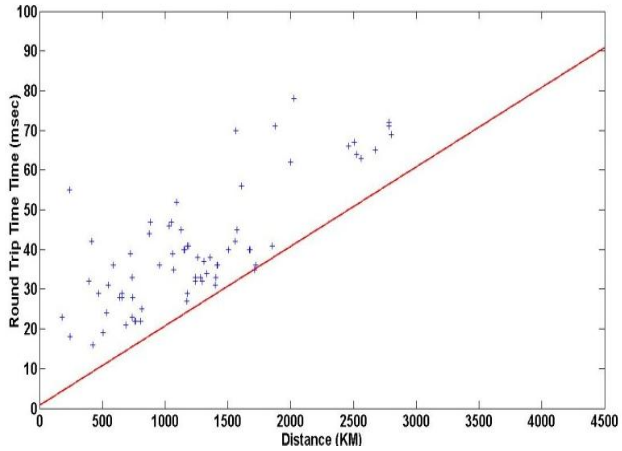


Figure 2. Example of best line computation in CBG.

This is a line which lies below all readings of round trip times on y axis and distance on x axis, so that it satisfies all of the distance-time relationship for that landmark host so it can incorporate any arbitrary host. The y-intercept of this line would represent the queueing delays. The slope should obey the rule of the thumb i.e. 0.1 ms round trip time per 100 km distance This will obviously lead to the most over-estimated distance possible for each landmark. But given that several landmarks are used, multilateration will reduce the error as the number of landmarks are increased.

Mathematically, the optimization problem can be written down as:

Minimize the objective function

$$\min_{\substack{m_i>0.01\\ b_i>0}} \sum_{i\neq j} y - m_i x - b_i \tag{2}$$

Subject to

$$y - m_i x - b_i > 0; \forall i \neq j \tag{3}$$

where y and x represent the round trip time measurements and distance measurements for each landmark host and i , j represent that measurements are pairwise between Landmarks i and j. We need to find for each landmark the slope of its best line m_i and the y intercept b_i. It is a convex optimization problem with no analytical solution. The objective function states that a best line is to be calculated which maintains the smallest distance to all the distance-time points while the constraints are that the line should lie below the points so as to satisfy all distance-time relationship for a particular landmark. The optimization problem is run for each landmark in turn to estimate its best line. For all such best line calculation two other further constraints that must be satisfied are that the estimate of constant router delays be greater than or equal to zero, and the propagation speed should be at least greater than the rule of the thumb i.e., 0.1 ms round trip time per 100 km distance. The solution to the optimization problem for each landmark is obviously revolving around the distance time measurements to each landmark to estimate speed of signal propagation and router delays.

When measurements have outliers, the solution will be biased by these outliers and generate sub-optimal results for the remainder of the measurements that are in general conformity of the distance-time relationship. Thus, the solution would generate circular radii that overestimate the host location by a large amount. The problem will not be severe if there a large number of landmarks with few landmarks having outlying or noisy distance-time relationships; but the problem becomes critical when the location error becomes dependent on few critical landmarks; and is worsened if they have outliers in measurements.

In this work we investigate the influence of the outliers on the solution of objective function minimization on the accuracy of location estimation. We see the how location error can be improved by satisfying maximum number of distance-time relationships in landmark best line computation after removing the outliers to get the tightest circles for better estimation of location of unknown host. In this work, we find that location error improvement can be of thousands of km. Our comparison is made with Gueye *et al*. [9] CBG technique.

4. Dealing With Outlier Detection And Removal

Statistical literature presents methods to deal with outliers in two ways, either remove them or mask their effects. The former would mean removal of outlying points completely before proceeding with the optimization problem for landmark best line computation and the latter would assign higher weightage to points that are in conformity with distance-time relationship in objective function minimization. For outlier removal, we may consider the following two techniques:

- *Naive: Convex Hull Removal*: In this naïve method, only the points at the periphery which represent the devious points are removed, and the remaining are used for the objective function minimization. This technique has also been considered by Wong et al. in their paper [24].
- *Local Correlation Based Integral (LOCI) based Outlier Detection and Removal*: Isolating data into two classes: normal and abnormal has been considered by researchers in the past as theOne Class Classification Problem: [11, 14, 15, 21]. We use Local Correlation based Integral (LOCI) method for the removal of anomalous measurements (outliers) so that the remaining measurements can be satisfied by the CBG of Gueye *et al*. [9] in a more befitting manner. This method is primarily motivated by Papadimitriou *et al*. [17]. To get a feel for the working of the algorithm, consider Figure 3. For every point, a region of radius ‘*r*’ is considered, all points lying in this radius ‘*r*’ and then further investigated for their ‘*αr*’ neighborhood where (*α*<1); in Figure 3, for instance the ‘*αr*’ neighborhoods of the point ‘*p*’ have values of one, two, two and four respectively. The standard deviation of variations in the neighborhood values determine if the original point is an outlier or not. For instance, if points within its radius ‘*r*’ have much variation of their own ‘αr’ neighborhoods, then the original points is classified as outlier. The process is repeated for every point with multiple variation or ‘*r*’ and ‘*α*’ values for multiscale resolution. The justification for this hosts located are similar geographical distances from another host are likely to follow same routing policies for routing.

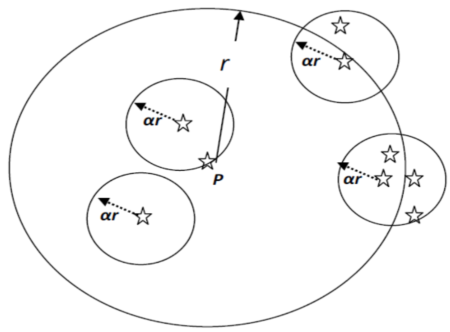


Figure 3. LOCI based outlier detection and removal [17].

To mask the effect of outlier we may consider an optimization based approach as below:

- *Iteratively Reweighted Least Squares* (IRLS): in Iteratively Reweighted Least Squares [7, 19], the objective function of the optimization problem is converted from L1 norm minimization of the error to weighted L2 norm minimization. Mathematically, the optimization problem is posed as:

Minimize the objective function,

$$\min_{\substack{m_i>0.01\\ b_i>0}} \sum_{i\neq j}[W.*(y-m_i x-b_i).*(y-m_i x-b_i)] \quad (4)$$

which is essentially the weighted L1 norm of the error vector.

Where, *W* is the vector of weights initially all elements are unity, (elements equal to number of distance-time points of landmark) and ‘.*’ denotes element wise multiplication.

Weights are updated as:

$$W_{new(i)} = \frac{W_{prev(i)}}{e_i} \quad (5)$$

Where,

$$\begin{aligned} Error &= [W.*(y-m_i x-b_i).*(y-m_i x-b_i)] \\ &= [e_1 \quad e_2 \quad \dots \quad e_n] \end{aligned} \quad (6)$$

The objective function is recursively computed until errors have converged.

5. Performance Evaluation

To evaluate the performance of the proposed framework, we only require the requisite end to end path metric and topology information. Throughout the remainder of this paper, we analyze the performance of overlay networks using real Internet datasets, so it is important that the methodology of obtaining this datasets is explicitly described before proceeding any further. Our datasets include a deployed Internet wide US based experimental network, Active Measurement Project (AMP) (*http://www.nlanr.net*), managed by National Laboratory for Applied Network Research (NLANR), which performs active measurements (delay measurements and trace routes) between 150 hosts in US and outside. From this set of AMP hosts we randomly selected 82 AMP. We obtained the latitude and longitudes of these 82 hosts using the publicly available utility (*www.latlong.net*). The distance computation between the hosts based on their latitude-longitude positions is by the following formula.

$$\begin{aligned}Distance(km) = 6378.135 * acos(cos(lat(host2)*pi/180)* \\ cos(lat(host1)*pi/180)* \\ cos(long(host1)*pi/180-long(host2)*pi/180) \\ +sin(lat(host2)*pi/180)* \\ sin(lat(host1)*pi/180));\end{aligned} \tag{7}$$

For the outlier rejection algorithms we use the Data Description Matlab Toolbox developed by Tax [22] with LOCI algorithm implementation by Janssens *et al.* [10, 11]. Figures 6, 7, 8, 9, and 10 show the location error improvement using the three strategies of outliers rejection and masking. In each case comparison of location error improvement is made by Gueye et al's CBG method [9]. Figure 4 shows the rejection of outliers at the boundary of the distance-time points through multiple stage naïve convex hull removal.

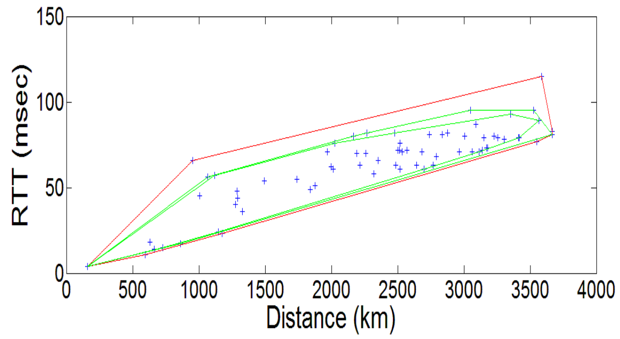


Figure 4. Outlier removal through naive convex hull removal (multi-stage).

Figure 5 shows the results when the very basic procedure is adopted for outlier rejection in the form of convex hull removal. When the convex hull removal is applied to all landmark data almost 50% of location error estimates are improved compared to basic CBG, the highest location improvement is of 1100 km; at the same time it degrades the location error improvement of theremaining 50% cases, the worst degradation is of 900km compared to CBG. Recursively applying thesame procedure only degrades the performance showing it is not effective in filtering outliers.

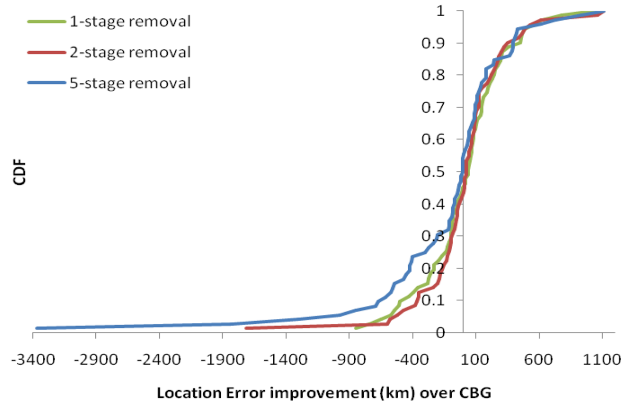


Figure 5. Location error improvement using naive convex hull removal for outliers (comparison is made with CBG).

Figure 6 shows the results when the very basic procedure is adopted for outlier masking in the form of a different objective function, Iteratively Reweighted Least Squares (lRLS). The results show that for up to 70% of the cases, IRLS is able to report better or similar results to CBG. The highest improvement is of 2500km.

As the final technique we consider the Local Correlation Integral (LOCI) method for outlier rejection. Figure 7 shows the typical action of LOCI technique in selection of outliers, a multi-resolution filtering of points who do not match the behavior of points in their neighborhood. The parameter k-sigma serves as a thresholding parameter to reject the deviant neighboring points, a lower value is a more stringent condition for rejecting deviant neighborhood points. The results shown in Figure 8, are much more impressive than Convex Hull removal considered before. Upto 75% of the hosts are located with lower error than CBG; the highest improvement being 2000km.

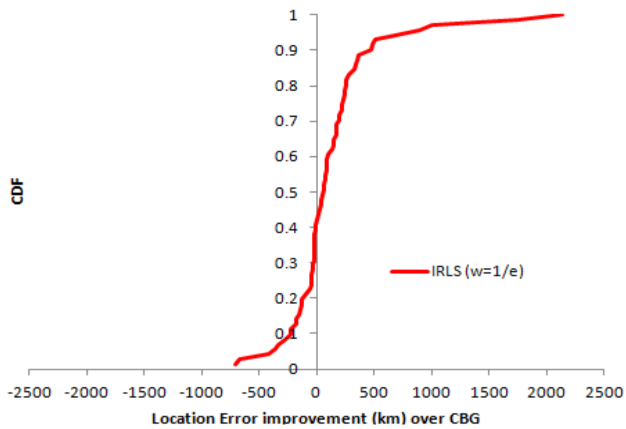


Figure 6. Location Error improvement by using iteratively reweighted least squares (comparison is made with CBG).

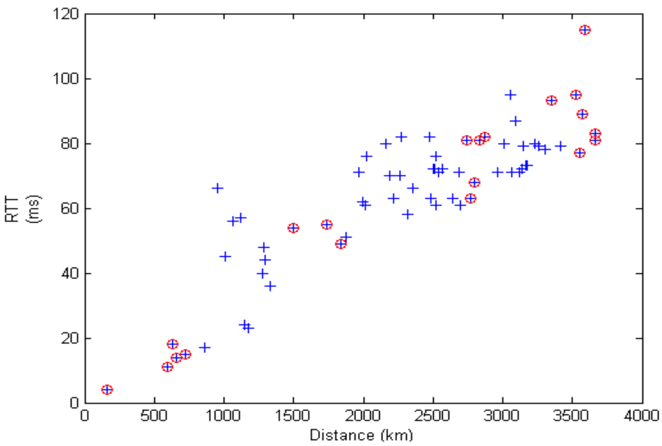


Figure 7. Outlier rejection shown in red circles via local correlation integral based method (ksigma=1.5).

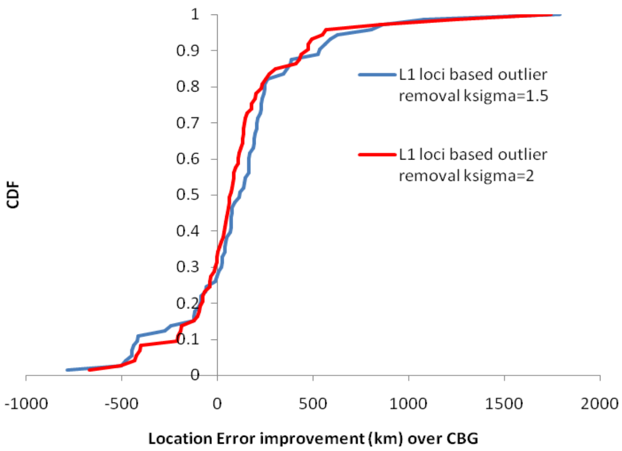


Figure 8. Location Error Improvement by using LOCI based outlier removal (comparison is made with CBG).

Figure 9 shows multilateration based approach for estimating the location of a host located inside UCLA (amp-ucla in our dataset).

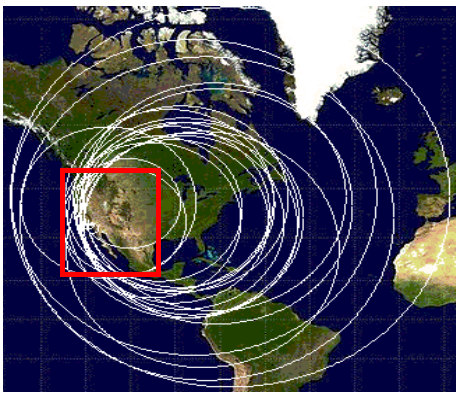
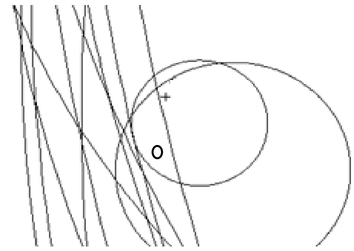


Figure 9. Practical multilateration based geoLocation of a host inside university of california los angeles (amp-ucla).

Figure 10 shows a zoomed view of this figure and it makes it clear that given the extreme coastal location of amp-ucla, most of the landmarks are located towards east of it. Thus a vast majority of the circles bounding the location of the host within a specific radius form the landmark host, tightens its location in a western direction. There are only two other landmark hosts (Figure 10), that tighten its position in the eastern and northern direction. Thus, the location error in this case is primarily dependent on the two smallest circles. The tightening of the radius of the landmark circles (corresponding to the two small circles especially) is crucial to lowering the geographical location estimation error of the unknown host (amp-ucla).

Figure 10 presents the difference in location error estimation when outlier detection is not used (Figure 10-a) and used (Figure 10-b). Notice the variation in the radii of the circles calculated for geolocating the unknown host. Although it appears that the radii of the larger circles are shown to be more relaxed (larger), the radii of the smaller circles are bounding the unknown host more tightly leading to reduction in its location estimate. In the first case the overlapping region has three vertices, the estimated location is the geometric center of triangle. However, in the second case, the overlapping common region has only two vertices, location estimate is just giving by the geometric center of the two intersection points of the common overlapping region. Location estimation error is 1813.6 km without outlier removal and just 69.1 km with outlier removal. Thus, Improvement in error is approx. 1744 km and results only due to the tightening of the two smaller circles.



a) Without outlier removal (top).

b) With outlier removal (bottom).

Figure 10. Zoomed view of Figure 9 (as depicted by rectangle). Plus symbol indicates true location and little circle represents estimated location.

Regarding the question of the time performance of the presented technique; Like in original CBG [9], considerable time may be required to solve the constraint based optimization problem in order to obtain the Best Line computation for each landmark. Considerable advances in the field of Genetic Algorithms (GA) have brought improvement to time performance of optimization problems [12]. However, this effort is only required once. Once the Best Lines are computed, the geometric approach for finding the centroid of the overlapping region of the circles in order to find the estimate of the geographic location of unknown host takes negligible time.

6. Conclusions

In this paper, we presented the challenges that arise in the optimization based approaches for geolocating Internet hosts. The optimization algorithms work best in optimization of mathematical parameters on the basis of real work physics, rather than understanding the policy-based routing model of the Internet. It is definitely important to incorporate such data so that the constraint based optimization models can produce more accurate estimates of the location. We also presented that optimization schemes which reduce the effects of outliers give more reasonable models and huge reduction of errors.

Refrences

[1] Andersen D., Balakrishnan H., Kaashoek F., and Morris R., "Resilient Overlay Networks," *ACM SIGOPS Operating Systems Review*, vol. 35, no. 5, p. 131-145, 2001.

[2] Augustin B., Friedman T., and Teixeira R., "Measuring Multipath Routing in the Internet," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 830-840, 2011.

[3] Chua D., Kolaczyk E., and Crovella M., "Network Kriging," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 12, pp. 2263-2272, 2006.

[4] Coates M., Pointurier Y., and Rabbat M., "Compressed Network Monitoring for ip and all-Optical Networks," *in Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, San Diego, pp. 241-252, 2007.

[5] Comarela G. and Crovella M., "Identifying and Analyzing High Impact Routing Events with PathMiner," *in Proceedings Internet Measurement Conference*, Vancouver, pp. 421-434, 2014.

[6] Dailey D. and Bell B., "A Method for GPS Positioning," *IEEE Transactions of Aerospace and Electronic Systems*, vol. 32, no. 3, pp. 1148-1154, 1996.

[7] Daubechies I., DeVore R., Fornasier M., and Gunturk S., "Iteratively Re-Weighted Least Squares Minimization: Proof of Faster than Linear Rate for Sparse Recovery," *in Proceedings of 42nd Annual Conference on Information Sciences and Systems*, Princeton, pp. 26-29, 2008.

[8] Dong Z., Perera R., Chandramouli R., and SubbalakshmiK P., "Network Measurement based Modeling and Optimization for IP Geolocation," *Computer Networks*, vol. 56, no. 1, pp. 85-98, 2012.

[9] Gueye B., Ziviani A., Crovella M., and Fdida S., "Constraint-Based Geolocation of Internet Hosts," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1219-1232, 2006.

[10] Janssens J., Postma E., and Herik J., "Density-Based Anomaly Detection in the Maritime Domain," *in Situation Awareness with Systems of Systems*, New York, pp. 119-131, 2013.

[11] Janssens J., Outlier Selection and One-Class Classification, Thesis, Tilburg University, 2013.

[12] Javidi M. and HosseinpourFard R., "Chaos Genetic Algorithm Instead Genetic Algorithm," *The International Arab Journal of Information Technology*, vol. 12, no. 2, pp. 163-168, 2015.

[13] Katz-Bassett E., John J., Krishnamurthy A., Wetherall D., Anderson T., and Chawathe Y., "Towards IP Geolocation using Delay and Topology Measurements," *in Proceedings Internet Measurement Conference*, Rio de Janeriro, pp. 71-84, 2006.

[14] Khan S. and Madden M., "One-Class Classification: Taxonomy of Study and Review of Techniques," *The Knowledge Engineering Review*, vol. 29, no. 3, pp. 345-374, 2014.

[15] Leng Q., Qi H., Miao J., Zhu W., and Su G., "One-Class Classification with Extreme Learning Machine," *Mathematical Problems in Engineering*, vol. 2015, pp. 1-11, 2015.

[16] Malaney R., "Securing Wi-Fi Networks with Position Verification: Extended Version," *International Journal of Security and Networks*, vol. 2, no. 1-2, pp. 27-36, 2007.

[17] Papadimitriou S., Kitagawa H., Gibbons P., and Faloutsos C., "LOCI: Fast Outlier Detection Using the Local Correlation Integral," *in Proceedings of the 19th International Conference on Data Engineering*, Bangalore, pp. 315-326, 2003.

[18] Pham N. and Pagh R., "A Near-Linear Time Approximation Algorithm for Angle-based Outlier Detection in High-dimensional Data," *in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, pp. 877-885, 2012.

[19] Qazi S. and Moors T., "On the Impact of Routing Matrix Inconsistencies on Statistical Path Monitoring in Overlay Networks," *Computer Networks*, vol. 54, no. 10, pp. 1554-1572, 2010.

[20] Qazi S. and Moors T., "Practical Issues of Statistical Path Monitoring in Overlay Networks with Large, Rank-Deficient Routing Matrices," *in Proceedings of 5th International Conference on Broadband Communications, Networks and Systems*, London, pp. 396-403, 2008.

[21] Tax D., One-class Classification, thesis, Delft University of Technology, 2001.

[22] Tax D., Data Description Tool Box for Matlab, PRTools, 2013.

[23] Wang Y., Kuzmanovic A., Burgener D., Huang C., and Flores M., "Towards Street-Level Client-Independent IP Geolocation," *in Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*, Boston, pp. 365-379, 2011.

[24] Wong B., Stoyanov I., and Sirer E., "Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts," *in Proceedings of the 4th USENIX conference on Networked Systems Design and Implementation*, Cambridge, pp. 313-326, 2007.

Sameer Qazi received his B.E. degree from National University of Sciences and Technology, Pakistan, in 2001 and the MS and PhD degrees from the University of New South Wales, Australia, in 2004 and 2009. He is currently working as Associate Professor in the College of Engineering at Karachi Institute of Economics and Technology, Pakistan. His research interests are Computer Network Optimization Problems and Cloud Computing.

Muhammad Kadri has completed his doctoral studies from University of Oxford, UK in 2009. He is currently working as Professor in the College of Engineering at Karachi Institute of Economics and Technology, Pakistan. His research interests include neuro-fuzzy control, system identification and optimization. He has published more than 40 papers in international journals and conferences