Constraint-Based Geolocation of Internet Hosts

Bamba Gueye
Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie
8, rue du Capitaine Scott
75015 - Paris - France
Bamba.Gueye@lip6.fr

Artur Ziviani
Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie
8, rue du Capitaine Scott
75015 - Paris - France
Artur.Ziviani@lip6.fr

Mark Crovella
Department of Computer Science.
Boston University
111 Cummington St – 02215
Boston, MA - USA
crovella@cs.bu.edu

Serge Fdida
Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie
8, rue du Capitaine Scott
75015 - Paris - France
Serge.Fdida@lip6.fr

ABSTRACT

Geolocation of Internet hosts enables a diverse and interesting new class of location-aware applications. Previous measurement-based approaches use reference hosts, called landmarks, with a well-known geographic location to provide the location estimation of a target host. This leads to a discrete space of answers, limiting the number of possible location estimates to the number of adopted landmarks. In contrast, we propose Constraint-Based Geolocation (CBG), which infers the geographic location of Internet hosts using multilateration with distance constraints, thus establishing a continuous space of answers instead of a discrete one. CBG accurately transforms delay measurements to geographic distance constraints, and then uses multilateration to infer the geolocation of the target host. Our experimental results show that CBG outperforms the previous measurement-based geolocation techniques. Moreover, in contrast to previous approaches, our method is able to assign a confidence region to each given location estimate. This allows a location-aware application to assess whether the location estimate is sufficiently accurate for its needs.

Categories and Subject Descriptors: C.2.4 [Computer-Communication Networks]: Distributed Systems

General Terms: Algorithm, Measurement

Keywords: Geolocation, Multilateration, Delay measurements.

1. INTRODUCTION

Novel location-aware applications could be enabled by an efficient means of inferring the geographic location of Internet hosts. Examples of such location-aware applications include targeted advertising on web pages, automatic selection of a language to display content, restricted content delivery following regional policies, and authorization of transactions only when performed from pre-established locations. Nevertheless, inferring the location of Internet hosts from their IP addresses is a challenging problem because there is no direct relationship between the IP address of a host and its geographic location. Previous work on the measurement-based geolocation of Internet hosts [6, 13] uses the positions of landmarks, reference hosts with well-known geographic location, as the possible location estimates for the target host. This leads to a discrete space of answers, *i.e.* the number of answers is equal to the number of reference hosts, that can be inaccurate because the closest reference host may still be far from the target.

To overcome this limitation, we propose the Constraint-Based Geolocation (CBG) approach, which infers the geographic location of Internet hosts using multilateration. Multilateration refers to the process of estimating a position using a sufficient number of distances to some fixed points. As a result, multilateration establishes a continuous space of answers instead of a discrete one. We use a set of landmarks to estimate the location of other Internet hosts. The fundamental idea is that given geographic distances to a given target host from the landmarks, an estimation of the location of the target host would be feasible using multilateration, just as the Global Positioning System (GPS) [4] does.

A key element of CBG is its ability to accurately transform delay measurements into distance constraints. The starting point is the fact that digital information travels along fiber optic cables at almost exactly $2/3$ the speed of light in a vacuum [7]. This means that any particular delay measurement immediately provides an *upper bound* on the great-circle distance between the endpoints. The upper bound is the delay measurement divided by the speed of light in fiber. Looking at this from the standpoint of a particular pair of endpoints, we can reason that there is some theoretical minimum delay

IMC'04, October 25–27, 2004, Taormina, Sicily, Italy.

for packet transmission that is dictated by the great-circle distance between them. Therefore, the actual measured delay between them involves only an *additive* distortion.

However, if CBG were to use simple delay measurements directly to infer distance constraints, it would not be very accurate. For accurate results, it is important to estimate and remove as much of the additive distortion as possible. CBG does this by self-calibrating the delay measurements taken from each measurement point. This is done in a distributed manner as explained in Section 3. After self-calibration, CBG can more accurately transform a set of measured delays to a target into distance constraints. CBG then uses multilateration with these distance constraints to establish a geographic region that contains the target host. Given the target region, a reasonable "guess" as to the host's location is at the region's centroid, which is what CBG uses as a point estimate of the target's position. Note that, in contrast to previous approaches, CBG is able to assign a confidence region to the given location estimate. This allows a location-aware application to assess whether the estimate is sufficiently accurate for its needs.

We evaluate CBG using real-life datasets with hosts that are geographically distributed through the continental U.S. and Western Europe. Our experimental results are promising and show that CBG outperforms the previous measurement-based geolocation techniques. The median error distance is below 25 km for the Western Europe dataset and below 100 km for the U.S. dataset. For the majority of evaluated target hosts, the obtained confidence regions allow a resolution at the regional level, *i.e.* about the size of a small U.S. state like Maryland or a small European country like Belgium.

This paper is organized as follows. Section 2 reviews the related work on this field and points out the contributions of CBG in contrast to previous approaches. In Section 3, we introduce CBG and its methodology to use multilateration with geographic distance constraints based on delay measurements to infer the location of Internet hosts. Following that, we present in Section 4 our experimental results. Finally, we conclude in Section 5.

2. GEOLOCATION OF INTERNET HOSTS

2.1 Related Work

A DNS-based approach to provide a geographic location service of Internet hosts is proposed in RFC 1876 [3]. Nevertheless, the adoption of the DNS-based approach has been limited since it requires changes in the DNS records and administrators have little motivation to register new location records. Tools such as IP2LL [9] and NetGeo [5] query Whois databases in order to obtain the location information recorded therein to infer the geographic location of a host.

Padmanabhan and Subramanian [6] investigate three different techniques to infer the geographic location of an Internet host. The first technique infers the location of a host based on the DNS name of the host or another nearby node. The second technique splits the IP address space into clusters such that all hosts with an IP address within a cluster are likely to be co-located. Knowing the location of some hosts in the cluster and assuming they are in agreement, the technique infers the location of the entire cluster. The third technique (GeoPing) is the closest to ours, as it is based on exploiting a possible correlation between geographic distance and network delay [6]. Given a set of landmarks with a well-known geographic location, the location estimate for a target host is the location of the landmark presenting the most similar delay pattern to the one observed for the target host.

In GeoPing, the number of possible location estimates is limited to the number of adopted landmarks, characterizing a discrete space of answers. In order to increase the accuracy of techniques like GeoPing, it is necessary to add additional landmarks. In Section 4.3, we compare CBG with GeoPing-like methods and show that CBG outperforms them.

2.2 Contributions

In this section, we summarize the contributions of CBG with respect to related work in geolocation of Internet hosts:

- CBG establishes a dynamic relationship between IP addresses and geographic location. This dynamic relationship results from a measurement-based approach where landmarks cooperate in a distributed and self-calibration manner, allowing CBG to adapt itself to time-varying network conditions. This contrasts with most previous work that relies on a static relationship;
- A major contribution of CBG is to point out that delay measurements can be transformed to geographic distance constraints to be used in multilateration. This potentially leads to more accurate location estimates of Internet hosts;
- CBG offers a continuous space of answers instead of a discrete one as do previous measurement-based approaches;
- CBG assigns a confidence region to each location estimate, allowing location-aware applications to assess whether the location estimate has enough resolution with respect to their needs.

3. CONSTRAINT-BASED GEOLOCATION

3.1 Multilateration with geographic distance constraints

The physical position of a given point can be estimated using a sufficient number of distances or angle measurements to some fixed points whose positions are known. When dealing with distances, this process is called multilateration. Similarly, when dealing with angles, it is called multiangulation. Strictly speaking, triangulation refers to an angle-based position estimation process with three reference points. However, quite often the same term is adopted for any distance or angle-based position estimation. In spite of the popularity of the term triangulation, we adopt the more precise term multilateration through the paper.

The main problem that stems from using multilateration is the accurate measurement of the distances between the target point to be located and the reference points. For example, the Global Positioning System (GPS) [4] uses multilateration to three satellites to estimate the position of a given GPS receiver. In the case of GPS, the distance between the GPS receiver and a satellite is measured by timing how long it takes for a signal sent from the satellite to arrive at the GPS receiver. Precise measurement of time and time interval is at the heart of GPS accuracy. In contrast to

GPS, it is a challenging problem to transform Internet delay measurements to geographic distances accurately. This is likely to be the reason why direct multilateration has remained so far unexploited for the purposes of geolocating Internet hosts. Hereafter, we explain the CBG design principles that enable the multilateration with geographic distance constraints.

For the location of Internet hosts using multilateration, we tackle the problem of estimating the geographic distance from the target host to be located to these landmarks given the delay measurements to the landmarks. The fundamental insight for the CBG methodology is that, no matter the reason, delay is only distorted additively with respect to the time for light in fiber to pass over the great-circle path. Therefore, we are interested in benefiting from this invariant by developing a method to estimate geographic distance *constraints* from these additively distorted delay measurements. How CBG use this insight to infer the geographic distance constraints between the landmarks and the target host from delay measurements is detailed in Section 3.2. It is also shown that as a consequence of the additive delay distortion, the resulting geographic distance constraints are generally overestimated with respect to the real distances.

3.2 From delay measurements to distance constraints

Before we introduce how CBG converts from delay measurements to geographic distance constraints, let us first observe a sample scatter plot relating geographic distance and network delay. This sample, shown in Fig. 1, is taken from the experiments described in Section 4. The x-axis is the geographic distance and the y-axis is the network delay between a given landmark L_i and the remaining landmarks. The meanings of "baseline" and "bestline" in Fig. 1 are explained along this section.

Recent work [6, 10, 13] investigates the correlation coefficient found within this kind of scatter plot, deriving a least squares fitting line to characterize the relationship between geographic distance and network delay. In contrast, we consider the *reasons* why points are scattered in the plot above, and argue that what is important is not the least-squares fit, but the tightest lower linear bound.

Based on these considerations, we propose a novel approach to establish a dynamic relationship between network delay and geographic distance. In order to illustrate this approach, suppose the existence of great-circle paths between the landmark L_i and each one of the remaining landmarks. Further, consider also that, when traveling on these great-circle paths, data are only subject to the propagation delay of the communication medium. In this perfect case, we should have a straight line comprising this relationship that is given by the slope-intercept form $y = mx + b$, where $b = 0$ since there are no localized delays and m is only related to the speed bits travel in the communication medium. As already noted, digital information travels along fiber optic cables at almost exactly 2/3 the speed of light in vacuum [7]. This gives a very convenient rule of 1 ms RTT per 100 km of cable. Such a relationship may be used to obtain an absolute physical lower bound on the RTT (or one-way delay) between sites whose geographic locations are well known. This lower bound is shown as the "baseline" in Fig. 1. In this idealized case, we could simply use this convenient rule to extract the accurate geographic distance between sites

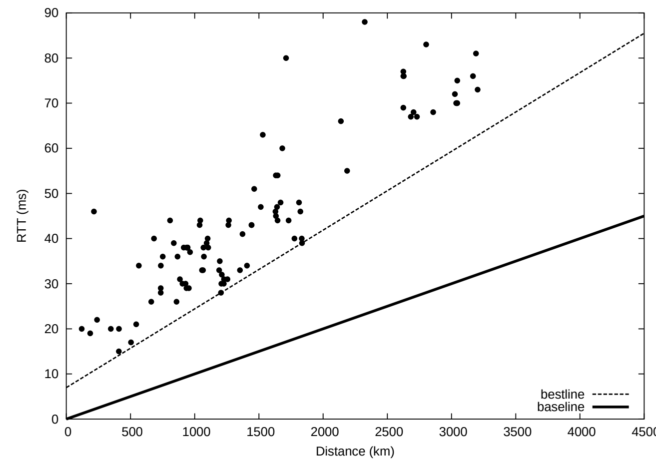


Figure 1: Sample scatter plot of geographic distance and network delay.

from delay measurements in a straightforward manner. Nevertheless, in practice, these great-circle paths rarely exist. Therefore, we have to deal with paths that deviate from this idealized model for several reasons, including queuing delay and lack of great-circle paths between hosts.

As stated in Section 3.1, the main insight behind CBG is that the combination of different sources of delay distortion with respect to the perfect great-circle case produces a pure geometric enhancement factor of the delay. We thus model the relationship between network delay and geographic distance using delay measurements in the following way. We define the "bestline" for a given landmark L_i as the line $y = m_i x + b_i$ that is closest to, but below, all data points (x, y) and has non-negative intercept, since it makes no sense to consider negative delays. Note that each landmark computes its own bestline with respect to all other landmarks. Therefore, the bestline can be seen as the line that captures the least distorted relationship between geographic distance and network delay from the viewpoint of each landmark.

The finding of the bestline is formulated as a linear programming problem. For a given landmark L_i, there are the network delay d_{ij} and the geographic distance g_{ij} toward each landmark L_j, where $i \neq j$. We need to find for each landmark L_i the slope m_i and the intercept b_i that determines the bestline given by the slope-intercept form $y = m_i x + b_i$. The condition that the bestline for each landmark L_i should lie below all data points (x, y) defines the feasible region where a solution should lie:

$$y - \frac{d_{ij} - b_i}{g_{ij}} x - b_i \geq 0, \quad \forall i \neq j, \tag{1}$$

where the slope $m_i = (d_{ij} - b_i)/g_{ij}$. The objective function to minimize the distance between the line with non-negative intercept and all the delay measurements is stated as

$$\min_{\substack{b_i \geq 0 \\ m_i \geq m}} \left(\sum_{i \neq j} y - \frac{d_{ij} - b_i}{g_{ij}} x - b_i \right), \tag{2}$$

where m is the slope of the baseline. Eq. (2) is used to find the solution m_i and b_i from Eq. (1) that determines the bestline for each landmark L_i.

Each landmark L_i then uses its own bestline to convert the delay measurement to the target host into a geographic distance. Thus, the estimated geographic distance constraint $\hat{g}_{i\tau}$ between a landmark L_i and the target host τ is derived from the delay distance $d_{i\tau}$ using the bestline of the landmark L_i as follows

$$\hat{g}_{i\tau} = \frac{d_{i\tau} - b_i}{m_i}. \quad (3)$$

If delays between landmarks are periodically gathered, this leads to a *self-calibrating* algorithm that determines how each landmark currently observes the dynamic relationship between network delay and geographic distance within the network.

3.3 Using distributed distance constraints to geolocate hosts

CBG uses a geometric approach using multilateration to estimate the location of a given target host τ. Each landmark L_i infers its geographic distance constraint to the target host τ, which is actually the additively distorted distance $\hat{g}_{i\tau} = g_{i\tau} + \gamma_{i\tau}$, using Eq. (3). Therefore, each landmark L_i estimates that the target host τ is somewhere within the circumference of a circle $\mathcal{C}_{i\tau}$ centered at the landmark L_i with a radius equal to the estimated geographic distance constraint $\hat{g}_{i\tau}$. Given K landmarks, the target host τ has a collection of closed curves $\mathbf{C}_\tau = \{\mathcal{C}_{1\tau}, \mathcal{C}_{2\tau}, \ldots, \mathcal{C}_{K\tau}\}$ that can be seen as an order-K Venn diagram. Out of the possible 2^K regions defined by this order-K Venn diagram for the target host τ, we are interested in the unique region $\mathcal{R}$ that forms the intersection of all closed curves $\mathcal{C}_{i\tau} \in \mathbf{C}_\tau$ given by

$$\mathcal{R} = \bigcap_{i}^{K} \mathcal{C}_{i\tau}. \quad (4)$$

Note that $\mathcal{R}$ is convex, since the regions $\mathcal{C}_{i\tau}$ are convex, and the intersection of convex sets is itself convex.

4. EXPERIMENTAL RESULTS

4.1 Datasets

- RIPE – data collected in the Test Traffic Measurements (TTM) project of the RIPE network [8]. The dataset we consider is composed by the 2.5 percentile of the one-way delay observed from each RIPE host to each other host in the set during a period of 10 weeks from early December 2002 until February 2003. Each RIPE host generates approximately 300 kB per day toward every other RIPE host with an average of two packets sent per minute. Most RIPE hosts are located in Europe and they are all equipped with GPS cards, thus allowing their exact geographic position to be known. We then use the 42 RIPE hosts located in Western Europe (W.E.) to compose our W.E. landmark dataset.

- NLANR AMP – data collected in the NLANR Active Measurement Project (AMP) [1]. The dataset we consider is composed by the 2.5 percentile of the RTT delay between all the participating nodes located in the continental United States (U.S.), in a total of 95 hosts. This data was collected on January 30, 2003 and is symmetric. Delay is sampled on average once a minute. This leads to an average measurement load of about 144 kB per day sent by each AMP host toward each other AMP host. The exact location of each participating node (in pairs of latitude and longitude) is also available. These 95 AMP hosts compose our U.S. landmark dataset.

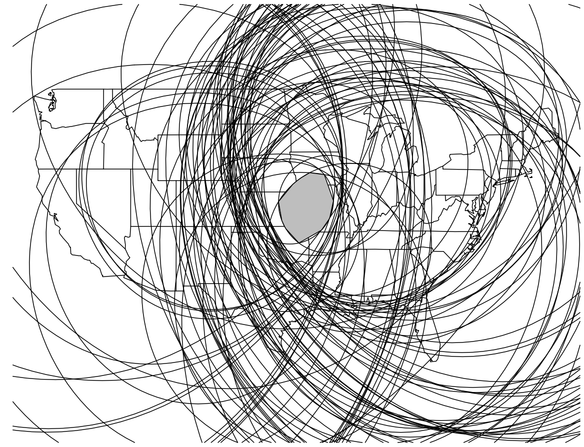


Figure 2: Location estimation of a target host.

In our experiments, the hosts in each dataset play one at a time the role of target host to be located. The remaining hosts in the same dataset are then considered as landmarks to perform the location estimation of the target host. The bestline of each landmark is computed using the set of landmarks of each scenario, thus excluding the target host. We repeat this procedure to evaluate the resulting location estimation of each host in both the U.S. and W.E. landmark datasets.

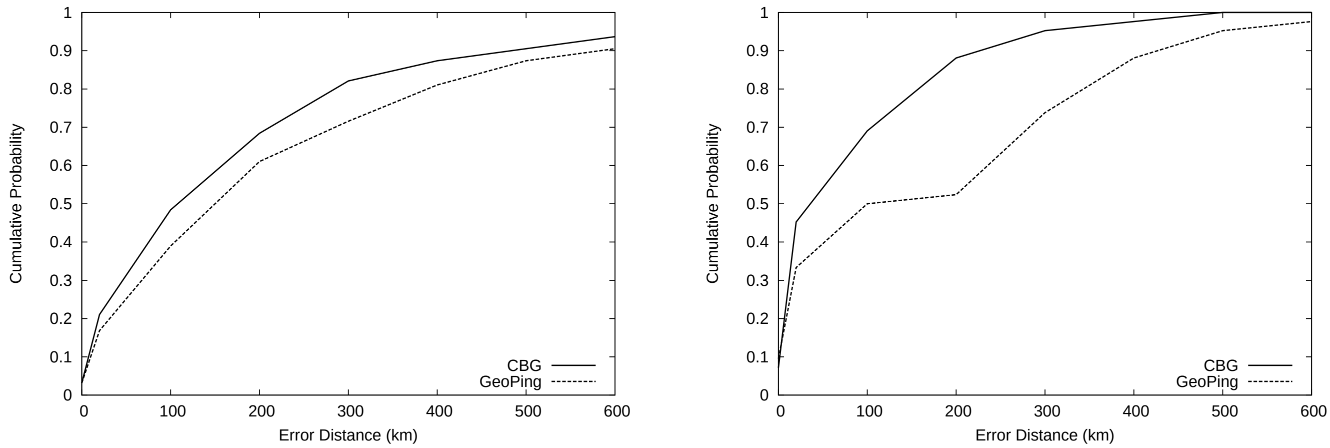
4.2 Location estimation of a target host

From the geographic distance constraints, CBG determines for each target host τ a set of closed curves $\mathbf{C}_\tau = \{\mathcal{C}_{1\tau}, \mathcal{C}_{2\tau}, \ldots, \mathcal{C}_{K\tau}\}$ (see Section 3.3), where K=42 for the W.E. dataset and K=95 for the U.S. dataset. Each curve in $\mathbf{C}_\tau$ is centered at its respective landmark L_i and has as radius the estimated geographic distance constraint $\hat{g}_{i\tau}$.

To illustrate the CBG methodology, Fig. 2 shows an example set of closed curves extracted from our experimental study. The area of the intersection region $\mathcal{R}$, *i.e.* the gray area in Fig. 2, indicates the confidence region that CBG associates with each location estimate. Note that in most cases confidence regions have a relatively small area, not visible in similar plots with all closed curves (Section 4.4 presents results on the sizes of confidence regions). This example has a larger confidence region than is typical, but is chosen so that the region is sufficiently visible so as to illustrate the CBG methodology.

4.3 Geolocating Internet hosts

The region $\mathcal{R}$ is the location estimate of CBG. Given this region, a reasonable "guess" as to the target host's location is at the region's centroid. Therefore, CBG uses the centroid



(a) U.S. dataset

(b) Western Europe dataset

Figure 3: Error distance for CBG and GeoPing.

of region $\mathcal{R}$ as a point estimate of the target's position.

We adopt the following heuristic to approximate the intersection region $\mathcal{R}$, *i.e.* the location estimate associated by CBG with the target host τ, by a polygon. The resulting polygon is used to approximately measure the area of the region $\mathcal{R}$ and provide an estimate of the point location of the target host. To form the polygon, we consider as vertices the crossing points of the circles $\mathcal{C}_{i\tau}$ that belong to all circles. Since the region $\mathcal{R}$ is convex, the polygon is an underestimate of the area of $\mathcal{R}$. We then approximate the region $\mathcal{R}$ by a polygon made up of line segments between N vertices $v_n = (x_n, y_n)$, $0 \leq n \leq N-1$. The last vertex $v_N = (x_N, y_N)$ is assumed to be the same as the first, *i.e.* the polygon is closed. These vertices of the polygon associated with a target host τ are the intersection points that belong to all circles $\mathcal{C}_{i\tau}$. The area of a non-self-intersecting polygon with vertices $v_0 = (x_0, y_0), \ldots, v_{N-1} = (x_{N-1}, y_{N-1})$ is given by

$$A = \frac{1}{2} \sum_{n=0}^{N-1} \begin{vmatrix} x_n & x_{n+1} \\ y_n & y_{n+1} \end{vmatrix} \tag{5}$$

where $|\mathbf{M}|$ denotes the determinant of matrix $\mathbf{M}$. The centroid c of the polygon, *i.e.* the position estimate of the target host τ, is positioned at (c_x, c_y) given by

$$c_x = \frac{1}{6A} \sum_{n=0}^{N-1} (x_n + x_{n+1}) \begin{vmatrix} x_n & x_{n+1} \\ y_n & y_{n+1} \end{vmatrix} \tag{6}$$

and

$$c_y = \frac{1}{6A} \sum_{n=0}^{N-1} (y_n + y_{n+1}) \begin{vmatrix} x_n & x_{n+1} \\ y_n & y_{n+1} \end{vmatrix}. \tag{7}$$

The point estimate of the target host and the estimate of the confidence region are the centroid (c_x, c_y) and the area A of the approximated polygon, respectively.

After inferring the point estimate for each considered target host, we compute the error distance, which is the difference between the estimated position and the real location of the target host τ. We compare our performance with the results obtained by a measurement-based geolocation system with a discrete space of answers [6, 13], *i.e.* where the location of the landmarks are used as location estimates. Fig. 3 shows the cumulative distribution function (CDF) of the observed error distance using CBG and an approach with a discrete set of answers like GeoPing. CBG outperforms the previous measurement-based discrete geolocation technique. The performance gap between the two approaches is more significant in the Western Europe dataset. This is probably because this dataset presents fewer landmarks than the U.S. dataset. In the discrete space approach, since the number of possible answer is limited to the locations of the landmarks, the number and placement of landmarks is a key point to the performance [12].

Considering the CBG results, the mean error distance in the U.S. dataset is 182 km, whereas for the W.E. dataset the mean error distance is 78 km. Most hosts in both landmark datasets have a quite good location estimation. The median error distance and the 80^{th} percentile for the U.S. dataset are 95 km and 277 km, respectively. In the W.E. dataset, the median error distance is 22 km and the 80^{th} percentile is 134 km.

4.4 Confidence region of a location estimation

The total area of the intersection region $\mathcal{R}$ is somewhat related to the confidence that CBG assigns to the resulting location estimate. Intuitively, this area quantifies the geographic extent or spread of each location estimate in km^2. The smaller the area of region $\mathcal{R}$, the more confident CBG is in this location estimate. Therefore, in contrast to previous measurement-based geolocation techniques, CBG assigns a confidence region in km^2 to each location estimate. We believe this is important because this confidence region may be used by location-aware applications to evaluate to which extent they can rely on the given location estimate. Furthermore, we envisage location-aware applications with

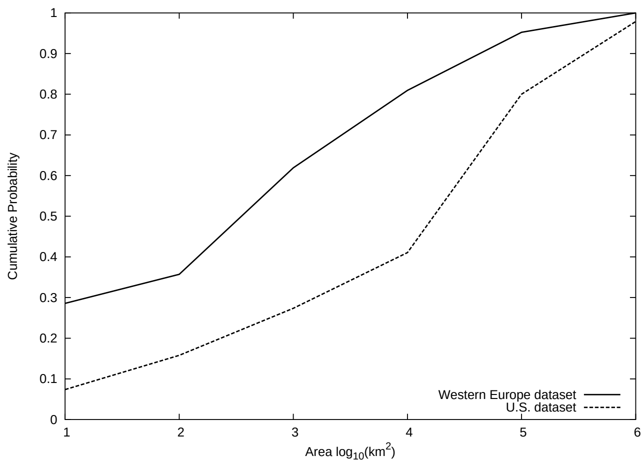


Figure 4: Confidence regions provided by CBG in km^2.

different requirements on accuracy. By using the confidence region, these location-aware applications may decide if the provided location estimate has sufficient resolution with respect to their particular needs.

Fig. 4 presents the CDF of the confidence regions in km^2 for the location estimates in both the U.S. and W.E. landmark datasets. Results show that, for the U.S. dataset, CBG assigns a confidence region with a total area less than 10^5 km^2 for around 80% of the location estimates. This area is slightly larger than Portugal or the U.S. state of Indiana. For the W.E. dataset, 80% of the location estimates have a confidence region of up to 10^4 km^2, thus enabling regional location. A confidence region of less than 10^3 km^2, which is equivalent to a large metropolitan area, is achieved by 25% of target hosts for the U.S. dataset and by 65% of target hosts for the W.E. dataset.

5. CONCLUSION

In this paper, we have proposed the Constraint-Based Geolocation (CBG), a measurement-based method to estimate the geographic location of Internet hosts. Based on delay measurements, CBG uses multilateration to infer a location estimate for a given target host. The accurate transformation of delay measurements to geographic distances is challenging because of many inherent characteristics of the current use and deployment of the Internet. Among these characteristics are queuing delays and the absence of great-circle paths between hosts. CBG contributes by pointing out that an accurate transformation from delay measurements to geographic distances *constraints* is indeed feasible. Moreover, CBG shows that in practice these constraints are often tight enough to allow an accurate location estimation using multilateration. CBG establishes a dynamic relationship between network delay and geographic distance. This is done in a distributed and self-calibrating fashion among the adopted landmarks using the bestline method.

Our experimental results show that CBG outperforms the previous measurement-based geolocation techniques. The median error distance obtained in our experiments for the U.S. dataset is below 100 km while for the Western Europe dataset this value is below 25 km. These results contrast with median error distances of about 150 km for the U.S. dataset and 100 km for the Western Europe dataset when GeoPing-like methods are used. Further, in contrast to previous approaches, CBG assigns a confidence region to each location estimate. This is important to allow a location-aware application to assess whether the location estimate is sufficiently accurate for its needs. Our findings indicate that an accurate location estimate, *i.e.* with a relatively small confidence region, is provided for most cases in both datasets, thus enabling location information at a regional level granularity. We mean by regional level the size of a small U.S. state or a small European country. It might be possible, once the confidence region has been determined, to use other methods if necessary to geolocate more precisely the target host using regional landmarks. This is left for future work.

Our results are based on measurements taken in well-connected, geographically contiguous networks. To some extent our work takes advantage of the fact that network connectivity has improved dramatically in the last decade, and that the relationship between network delay and geographic distance is strong in these regions [2, 11, 13]. Location to or from typical end-systems is part of our future work. Thus one must be cautious before extrapolating our present results to arbitrary network regions.

6. REFERENCES

[1] *NLANR Active Measurement Project.* `http://amp.nlanr.net/active/`.

[2] K. Claffy. Internet measurement: myths about Internet data. Talk at NANOG24 Meeting, Feb. 2002. `http://www.caida.org/outreach/presentations/Myths2002/`.

[3] C. Davis, P. Vixie, T. Goowin, and I. Dickinson. A means for expressing location information in the domain name system. *Internet RFC 1876*, Jan. 1996.

[4] P. Enge and P. Misra. Special issue on global positioning system. *Proceedings of the IEEE*, 87(1):3–15, Jan. 1999.

[5] D. Moore, R. Periakaruppan, J. Donohoe, and K. Claffy. Where in the world is netgeo.caida.org? In *Proc. of the INET'2000*, Yokohama, Japan, July 2000.

[6] V. N. Padmanabhan and L. Subramanian. An investigation of geographic mapping techniques for Internet hosts. In *Proc. of the ACM SIGCOMM'2001*, San Diego, CA, USA, Aug. 2001.

[7] R. Percacci and A. Vespignani. Scale-free behavior of the Internet global performance. *The European Physical Journal B - Condensed Matter*, 32(4):411–414, Apr. 2003.

[8] *RIPE Test Traffic Measurements.* `http://www.ripe.net/ttm/`.

[9] University of Illinois at Urbana-Champaign. *IP Address to Latitude/Longitude.* `http://cello.cs.uiuc.edu/cgi-bin/slamm/ip2ll/`.

[10] S. van Langen, X. Zhou, and P. van Mieghem. On the estimation of Internet distances using landmarks. In *Proc. of the International Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking – NEW2AN'04*, St. Petersburg, Russia, Feb. 2004.

[11] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the Internet's large-scale topology. *Proc. of the National Academy of Sciences (PNAS)*, 99:13382–13386, Oct. 2002.

[12] A. Ziviani, S. Fdida, J. F. de Rezende, and O. C. M. B. Duarte. Demographic placement for Internet host location. In *Proc. of the IEEE GLOBECOM'2003*, San Francisco, CA, USA, Dec. 2003.

[13] A. Ziviani, S. Fdida, J. F. de Rezende, and O. C. M. B. Duarte. Toward a measurement-based geographic location service. In *Proc. of the Passive and Active Measurement Workshop - PAM'2004*, Lecture Notes in Computer Science (LNCS) 3015, pages 43–52, Antibes Juan-les-Pins, France, Apr. 2004.