

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221198599>

# Leveraging Buffering Delay Estimation for Geolocation of Internet Hosts

Conference Paper in Lecture Notes in Computer Science · May 2006

DOI: 10.1007/11753810\_27 · Source: DBLP

## CITATIONS

15

## READS

120

### 4 authors:



**Bamba Gueye**

Cheikh Anta Diop University, Dakar

44 PUBLICATIONS 594 CITATIONS

[SEE PROFILE](#)



**Steve Uhlig**

Queen Mary, University of London

199 PUBLICATIONS 5,791 CITATIONS

[SEE PROFILE](#)



**Artur Ziviani**

Laboratório Nacional de Computação Científica

181 PUBLICATIONS 1,747 CITATIONS

[SEE PROFILE](#)



**Serge Fdida**

Sorbonne Université

276 PUBLICATIONS 3,558 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



FIBRE project [View project](#)



ANA Project [View project](#)

# Leveraging Buffering Delay Estimation for Geolocation of Internet Hosts

Bamba Gueye<sup>1</sup>, Steve Uhlig<sup>2\*</sup>, Artur Ziviani<sup>3</sup>, Serge Fdida<sup>1</sup>

<sup>1</sup> Université Pierre et Marie Curie  
Laboratoire d'Informatique de Paris 6 (LIP6)  
{gueye, fdida}@rp.lip6.fr

<sup>2</sup> Université Catholique de Louvain  
Department of Computing Science and Engineering  
suh@info.ucl.ac.be

<sup>3</sup> National Laboratory for Scientific Computing (LNCC)  
ziviani@lncc.br

**Abstract.** Geolocation techniques aim at determining the geographic location of an Internet host based on its IP address. Currently, measurement-based geolocation techniques disregard the buffering delays that may be introduced at each hop along the path taken by probe packets. To fill this gap, we propose the *GeoBuD* (Geolocation using **B**uffering **D**elay estimation) approach. Although the network delay and the geographic distance between two Internet hosts have been shown to be related to some extent, leveraging buffering delay estimation at each hop for geolocation purposes is challenging for two reasons. First, correctly estimating the buffering delay at intermediate hops along a traceroute path for geolocation purposes depends on the accurate estimation of the geolocation of the intermediate routers. Second, even given an *a priori* knowledge of the location of the routers, estimating the buffering delays is difficult due to the coarse-grained information provided by delay measurements. Relying on traceroute measurements, we show that leveraging buffering delay estimation improves accuracy in the measurement-based geolocation of Internet hosts as well as the confidence that the geolocation service associates to each estimation.

**Keywords:** geolocation, buffering delay estimation, traceroute, multilateration.

## 1 Introduction

Geographically locating an Internet host from its IP address enables a diversified class of location-aware applications [1,2,3,4,5,6,7]. Examples of such applications comprise targeted advertising on web pages, displaying local events and regional weather, automatic selection of a language to first display the content of web pages, restricted content delivery following regional policies, authorization of transactions only when performed from pre-established locations, or locating pedo-criminality. Each application may have a different requirement on the resolution of the location estimation. Nevertheless, as IP addresses are in general allocated in an arbitrary fashion, there is no

---

\* Steve Uhlig is Postdoctoral fellow of the Belgian National Fund for Scientific Research (F.N.R.S.).

inherent relation between an IP address and the physical location of the corresponding physical interface. Therefore, inferring the geographic location of Internet hosts is a challenging problem.

Previous work on measurement-based geographic location of Internet hosts [8,9,10] relies on delay measurements between *landmarks*, *i.e.* hosts with well-known geographic location, to provide the position of a target host. In GeoPing [8], the positions of landmarks are used as the possible location estimates for a given target host. This leads to a discrete space of answers that may limit location accuracy because of the system's dependence on the number and placement of landmarks [9]. The *Constraint-Based Geolocation* (CBG) approach proposed by Gueye *et al.* [10] transforms delay measurements into distance constraints and then uses *multilateration* to estimate the geographic location of a given target host. Multilateration refers to the process of estimating a position using a sufficient number of distances to some fixed points, thus establishing a continuous space of answers instead of a discrete one. This multilateration with distance constraints provides an overestimation of the distance from each landmark to the target host to be located, thus determining a region that hopefully encloses the location of the target host. The centroid of this region is the location estimation provided by CBG. Further, the area size of this region is a confidence measure CBG associates with each given location estimation; the smaller the region, the more confident the system is in the provided estimation.

Although showing relatively accurate results in most cases, these measurement-based approaches may have their accuracy disturbed by many sources of distortion that affect delay measurements. For example, delay distortion may be introduced by the circuitous Internet paths that tend to unnecessarily inflate the end-to-end delay [11,12,13]. Another source of distortion is the unpredictable buffering delay that packets face in queues at the intermediate routers along the end-to-end path. For an accurate geolocation of Internet hosts based on delay measurements, it is crucial to estimate and remove as much of the additional delay as possible.

This paper investigates the distortion introduced in delay measurements due to buffering delays and possible counter-measures to it in order to improve geolocation techniques. We present GeoBuD, a novel way of geolocating Internet hosts. We rely on traceroute measurements to estimate the buffering delay introduced along the path from each landmark to the target. Based on traceroute information about the successive RTTs at each intermediate hop, we estimate the buffering delay introduced by each of these hops. Our results show that the estimation of buffering delays introduced along the path allows the improvement of the geolocation estimation given by CBG. This is so because the additional delay distortions caused by buffering delay are removed from the overestimations of distance constraints that define the region enclosing the target host in CBG, thus allowing tighter overestimations that result in a smaller region. Smaller regions that still enclose the target host provide more accurate location estimation in CBG.

The remainder of the paper is structured as follows. Section 2 discusses the related work. Section 3 describes the CBG approach to estimate the geographic location of a given target host. Section 4 explains our methodology for performing the traceroutes and estimating the buffering delays along the traceroute measurements. Section 5

compares the results of GeoBuD and those of the CBG approach. Finally, Section 6 concludes our paper and discusses future work.

## 2 Related work

A DNS-based approach to provide a geographic location service of Internet hosts is proposed in RFC 1876 [14]. 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 [15] and NetGeo [16] query Whois databases in order to obtain the location information recorded therein to infer the geographic location of a host. This information, however, may be inaccurate or stale. Moreover, if a large and geographically dispersed block of IP addresses [17] is allocated to a single entity, the Whois databases may contain just a single entry for the entire block.

There are also some geolocation services based on an exhaustive tabulation between IP addresses ranges and their corresponding locations. Examples of such services are GeoURL [1], the Net World Map project [2], and several commercial tools [3,4,5,6,7].

Padmanabhan and Subramanian [8] 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. This technique is the basis of GeoTrack [8], Visual-Route [18], GTrace [19], and the SarangWorld Traceroute project [20]. Quite often network operators assign names to routers that have some geographic meaning, presumably for administrative convenience. Nevertheless, not all names contain an indication of location. Since there is no standard, operators commonly develop their own rules for naming their routers even if the names are geographically meaningful. Therefore, the parsing rules to recognize a location from a node name must be specific to each operator. The creation and management of such rules is a challenging task as there is no standard to follow.
- 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. An example of such a technique is GeoCluster [8]. This technique, however, relies on information that is partial and possibly inaccurate. The information is partial because it comprises location information for a relatively small subset of the IP address space. Moreover, such information may be inaccurate because the databases rely on data provided by users, which may be unreliable.
- The third technique, GeoPing [8], is based on exploiting a possible correlation between geographic distance and network delay. The location estimation of a host is based on the assumption that hosts with similar network delays to some fixed probe machines tend to be located near each other. This assumption is similar to the one exploited by wireless positioning systems such as RADAR [21] concerning the relationship between signal strength and distance. Therefore, given a set of landmarks with a well-known geographic location, the location estimation 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. As a consequence, the accuracy of this discrete space system is directly related to the number and placement of the adopted landmarks [9].

To overcome the limitation of using a discrete space of answers, the Constraint-Based Geolocation [10] approach uses multilateration to yield a continuous space of answers. In the next section, we provide a brief background on how the CBG methodology operates as the goal in this paper is to investigate the impact that the leveraging of buffering delay can have on the geolocation of Internet hosts based on multilateration.

### 3 Background on the CBG approach

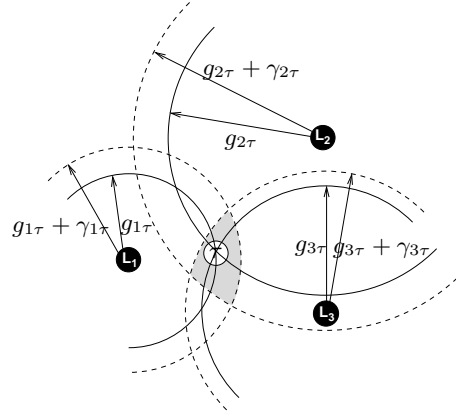
In this section, we present a brief background on how CBG provides geolocation estimation for target hosts based on delay measurements.

#### 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.

Consider a set  $\mathcal{L} = \{L_1, L_2, \dots, L_K\}$  of  $K$  landmarks. Landmarks are reference hosts with a well-known geographic location. For the location of Internet hosts using multilateration, CBG [10] tackles the problem of estimating the geographic distance from these landmarks towards the target host to be located, given the delay measurements from the landmarks. From a measurement viewpoint, the end-to-end delay over a fixed path can be split into two components: a deterministic (or fixed) delay and a stochastic delay [22]. The deterministic delay is composed by the minimum processing time at each router, the transmission delay, and the propagation delay. This deterministic delay is fixed for any given path. The stochastic delay comprises the queuing delay at the intermediate routers and the variable processing time at each router that exceeds the minimum processing time. Besides the stochastic delay, the conversion from delay measurements to geographic distance is also distorted by other sources as well, such as circuitous routing and the presence of redundant data. Anyway, it should be noted that no matter the source of distortion, this delay distortion is always additive with respect to the minimum delay of an idealized direct great-circle path.

Figure 1 illustrates the multilateration in CBG using the set of landmarks  $\mathcal{L} = \{L_1, L_2, L_3\}$  in the presence of some additive distance distortion due to imperfect measurements. Each landmark  $L_i$  intends to infer its geographic distance constraint to a target host  $\tau$  with unknown geographic location. Nevertheless, the inferred geographic distance constraint is actually given by  $\hat{g}_{i\tau} = g_{i\tau} + \gamma_{i\tau}$ , *i.e.* the real geographic distance  $g_{i\tau}$  plus an additive geographic distance distortion represented by  $\gamma_{i\tau}$ . This purely additive distance distortion  $\gamma_{i\tau}$  results from the possible presence of some additive delay distortion. As a consequence of having additive distance distortion, the location



**Fig. 1.** Multilateration with geographic distance constraints.

estimation of the target host  $\tau$  should lie somewhere within the gray area (cf. Figure 1) that corresponds to the intersection of the overestimated geographic distance constraints from the landmarks to the target host.

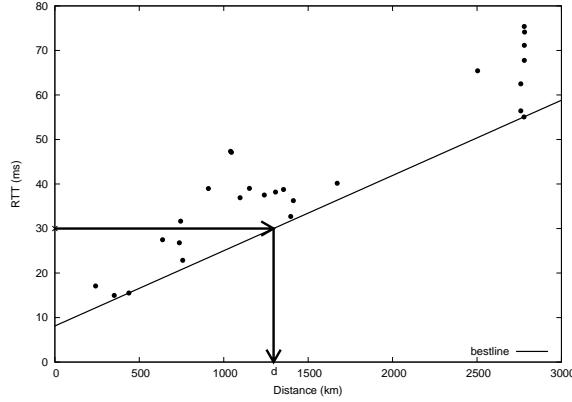
### 3.2 From delay measurements to distance constraints

Recent work [23,8,24] has investigated the correlation between geographic distance and network delay. Figure 2 provides an example of the relation between the distance and the delay for one of the landmarks we used in our measurements towards the remaining landmarks of our dataset (further details on the experimental data used are found in Section 5). The *bestline* shown in Figure 2 for a given landmark  $L_i$  is defined as the line that is closest to, but below all data points  $(x, y)$ , where  $x$  expresses the actual great-circle geographic distance between this given landmark and all the other landmarks in the set, while  $y$  represents the measured RTT between the same pairs. The equation of the bestline is defined as

$$y = m_i x + b_i. \quad (1)$$

It should be noted that each landmark finds its slope  $m_i$  and its positive intercept  $b_i$  based only on delay measurements between the available landmarks. For further details about the computation of  $b_i$  and  $m_i$ , we refer the reader to [10]. The presence of a positive intercept  $b_i$  in the bestline reflects the presence of some localized delay. Each landmark uses its own bestline to convert the delay measurement towards the target host into a geographic distance constraint. A delay measurement from the considered landmark of Figure 2 towards a particular target host  $\tau$  is transformed into a distance constraint by projecting the measured delay on the distance axis using the computed bestline of this landmark. For example, if the measured delay is 30 ms, the distance constraint is  $d$ , as illustrated by the thick arrow in Figure 2. This estimated geographic distance constraint  $\hat{g}_{i\tau}$  between a landmark  $L_i$  and a target host  $\tau$  is derived from the delay  $d_{i\tau}$  using the bestline of the landmark as follows:

$$\hat{g}_{i\tau} = \frac{d_{i\tau} - b_i}{m_i}. \quad (2)$$



**Fig. 2.** Sample scatter plot of geographic distance and network delay.

Each landmark  $L_i$  localizes a given destination  $\tau$  inside a circle whose radius is the obtained distance constraint  $\hat{g}_{i\tau}$ . The region formed by the intersection of all these circles from the set of landmarks is called in CBG the *confidence region*. CBG provides the centroid of this confidence region as the location estimation for the target host.

#### 4 Buffering delay estimation via traceroutes

CBG builds its distance constraints on a per-landmark basis. In contrast, based on a per-destination, GeoBuD transforms delay measurements into distance constraints. In practice, paths for different destinations may suffer from different distortions. To take that into account, we first replace the linear model of CBG by decomposing it in a per-hop basis. So in the GeoBuD approach, for each landmark  $L_i$  and each target host  $\tau$ , we model the delay  $y_{i\tau}$  as

$$y_{i\tau} = m_i x_{i\tau} + b_{i\tau}, \quad (3)$$

where  $m_i$  represents the propagation speed of data along the path computed only between the landmarks,  $x_{i\tau}$  represents the geographic distance constraint between landmark  $L_i$  and destination  $\tau$ , and  $b_{i\tau}$  represents the total buffering delay along the path from  $L_i$  to a target host  $\tau$ . In our measurements, the value of  $m_i$  actually represents 2 times the propagation speed of light in fiber, as  $m_i$  captures both the signal propagation aspect and the fact that the delay on which we are relying is a RTT hence contains both the forward and the return path. To estimate the total buffering delay  $b_{i\tau}$ , we estimate the buffering delay at each hop along the path from  $L_i$  to target  $\tau$  based on traceroute between these nodes. The output of the traceroute measurements hopefully provides the different intermediate nodes that compose the path, as well as the delay between each pair of consecutive intermediate router.

For example, suppose that we perform a traceroute from landmark  $L_i$  towards some target host  $\tau$ . The traceroute is composed of  $n$  intermediate hops, the last hop being the one that arrives at the target node. For each hop  $k$  of the traceroute that answers with an ICMP message `TIME exceeded`, we have an RTT measurement. If by any reason an intermediate router along the traceroute path does not answer with the ICMP message

TIME exceeded, we disconsider this hop as we lack a delay measurement for this particular node. To estimate  $b_{i\tau}$  in Equation (3), we actually estimate its components  $b_k$  along the traceroute path using

$$\Delta RTT_{k+1} = RTT_{k+1} - RTT_k = m_i \times dist(k, k+1) + b_{k+1}, \quad (4)$$

where  $k$  represents the  $k^{th}$  intermediate router on the traceroute path for which we were able to have a delay measurement and geographical location. The term  $RTT_k$  denotes the minimum RTT value out of the 3 RTTs measurements obtained for a given hop<sup>1</sup>  $k$  and  $dist(k, k+1)$  represents the geographic distance between nodes  $k$  and  $k+1$ . Note that  $m_i$  is the same as the one in Equation (1). The sum of the  $dist(k, k+1)$  for each  $k$  from 0 to  $n-1$  gives the estimation of the geographic path length followed by the traceroute. Thus, we estimate the buffering delay  $b_k$  at each hop  $k$  in a straightforward way from Equation (4) as

$$b_k = \Delta RTT_k - m_i \times dist(k-1, k). \quad (5)$$

It is clear from Equation (5) that we need to estimate the geographic distance between each pair of consecutive intermediate routers along the traceroute path in order to be able to estimate  $b_k$ . This implies knowing the geographic location of these routers. It is unlikely to have an *a priori* knowledge of the geographic location of all possible intermediate routers along a traceroute path. This would actually amount to being able to geolocate any Internet node, *i.e.* the actual intent of the geolocation service under investigation. Therefore, the estimation of buffering delay along the path demands successive use of the geolocation service on each node identified along the traceroute path until reaching the target host.

## 5 GeoBuD evaluation

In order to estimate the buffering delay  $b_k$  and the  $dist(k, k+1)$  at each hop  $k$  along a path between each landmark and each target host  $\tau$ , we have considered two datasets: First, we considered nodes located in the U.S. We have used 29 PlanetLab nodes [26] as landmarks and 87 AMP nodes [27] as targets. The dataset we consider is composed by traceroute measurements performed on October 17<sup>th</sup> 2005 from our landmarks towards targets hosts.

For the second dataset we performed on November 21<sup>st</sup> 2005 traceroutes from 27 PlanetLab nodes located in Western Europe towards 57 RIPE nodes [28], also located in Western Europe. In CBG methodology, landmarks perform ping measurements towards a given target host to locate it. Traceroute measurements from the same landmarks towards the same targets were performed simultaneously with ping measurements, in order to have similar network conditions for both CBG and traceroute measurements.

In our traceroute experiments, for the U.S. dataset, we have been able to geolocate 1153 distinct intermediate routers excluding the AMP hosts, out of a total of 1408 traversed routers, thus leading to geolocating 82% of the intermediate routers. In the W.E

---

<sup>1</sup> Each step of a traceroute consists in sending 3 consecutive UDP packets towards the destination using an increasing TTL value. In our measurements we rely on native traceroutes [25].



dataset we have located 1235 routers among the 1328 routers that we have encountered. So 93% of W.E. routers are located. It should be noted that most of undiscovered routers are typically located in the vicinity of the source or the destination, so the resulting error in the traceroute path length estimation is due to be small. For each of these located routers, we relied on the CBG-based GeoLIM project [29] to find out their geographic location. We cross-checked the results obtained with the GeoLIM project with *rockettrace* provided by the scriptroute tool suite [30]. In addition, some hops along the traceroutes underwent congestion by the time the measurements are carried. These traffic conditions, however, may not last for the whole time of the traceroute measurements. In such a case, only a few intermediate hops along the traceroute path exhibit a very large RTT value. If any of the intermediate hops along a traceroute exhibits a RTT value larger than any RTT of its succeeding hops along the traceroute path, we disconsider this particular hop. If we were to take such inflated RTTs into account we would overestimate the buffering delay for that hop.

After performing the geolocation of the intermediate routers, we compute the set of  $b_k$  values along each of different traceroute paths for the located intermediate nodes using Equation (4). In some cases, the estimated  $b_k$  are negative, in which case they were not considered. We had 21% of negative  $b_k$ 's corresponding to 4043 among 19172  $b_k$ 's that we have computed for AMP hosts. For RIPE hosts the percentage of negative  $b_k$ 's is 14% for 11908  $b_k$ 's found. Most cases where the  $b_k$  were negative correspond to situations where  $\Delta RTT_{k+1}$  is very small or negative. This is due to variations in the network conditions along the path of the traceroute during the experiments. Hence, to consider a particular  $b_k$ , we require that  $\Delta RTT_{k+1} > 0$ .

For each landmark  $L_i$  and target  $\tau$ , we have then a corresponding  $b_{i\tau} = \sum_{k=1}^{n-1} b_k$ , where  $n$  denotes the number of intermediate hops along the traceroute path from landmark  $L_i$  and target  $\tau$ . To transform delay measurements into distance constraints, we can use the following equation derived from Equation (3):

$$x_{i\tau} = \frac{y_{i\tau} - b_{i\tau}}{m_i}. \quad (6)$$

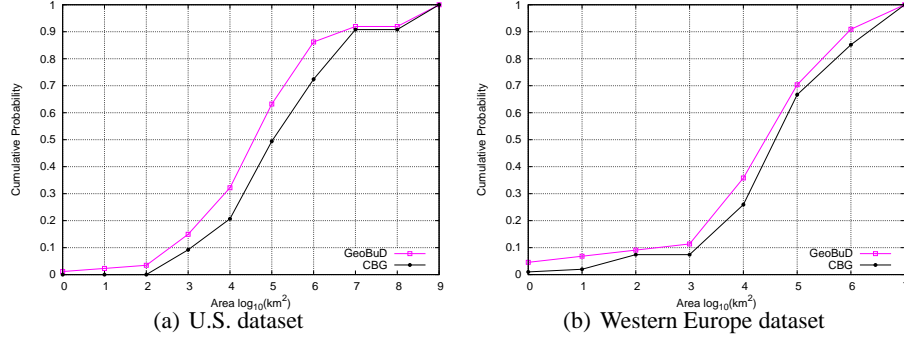
GeoBuD uses the distance constraints given by Equation (6) to localize a given target host. The distance constraints obtained by GeoBuD are expected to be tighter than those provided by the CBG method. Using these new tighter distance constraints, in spite of the number of negative  $b_k$ 's, the confidence region shrinks, thus increasing both the accuracy of the location estimation and the system's confidence on these location estimation as shown in Section 5.1.

## 5.1 Shrinking the confidence region

Figure 3 compares the cumulative probability distributions of the confidence region of GeoBuD and the CBG approach. On the  $x$ -axis, we have the surface area of the confidence region for different target hosts. On the  $y$ -axis, we show the probability that the location estimation for the target hosts have a confidence region smaller than  $x$ .

One can observe on Figure 3 the improvement due to buffering delay estimation for areas smaller than  $10^7$  km<sup>2</sup>. With CBG, 72% of the target hosts located in the U.S. have

a confidence region smaller than  $10^6 \text{ km}^2$ . For GeoBuD and the same confidence region surface, we have about 86% of the target hosts. With CBG, 49% of the target hosts have a confidence region smaller than  $10^5 \text{ km}^2$ , whereas for GeoBuD 63% of the target hosts are within such a confidence region. For hosts located in W.E., GeoBuD localizes 10% of the target hosts with a confidence region inferior to  $10^2 \text{ km}^2$ . For reference, a surface area of  $10^5 \text{ km}^2$  is slightly larger than Portugal or the U.S. state of Indiana.



**Fig. 3.** Confidence regions provided by GeoBuD and CBG in  $\text{km}^2$ .

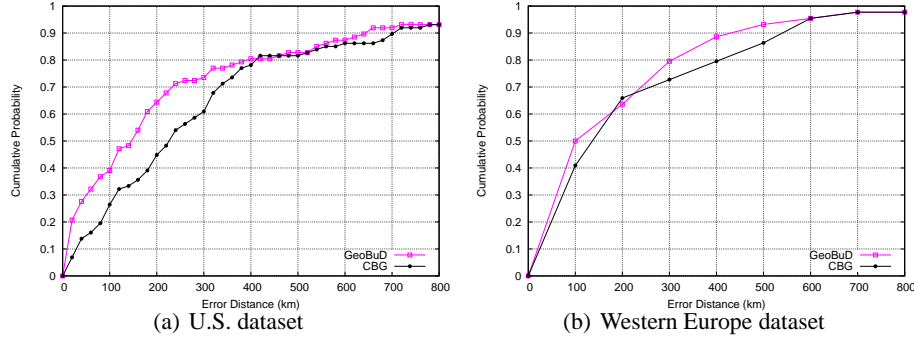
## 5.2 Location estimation error

We might expect that reducing the surface area of the confidence region by estimating the buffering delay would also reduce the error observed in location estimation. In Figure 4, we show the cumulative probability of the error observed in the obtained location estimation. The estimation error for a given target host is the difference between its actual geographic location and its location estimate. The performance gap between GeoBuD and CBG is larger in the U.S. dataset. For the U.S. dataset, 80% of the target hosts, the estimation error is smaller using GeoBuD compared to CBG. The median of the location estimation error is of 144 km for GeoBuD, while of 228 km for CBG. In the W.E. dataset, it is 100 km and 137 km for GeoBuD and CBG respectively.

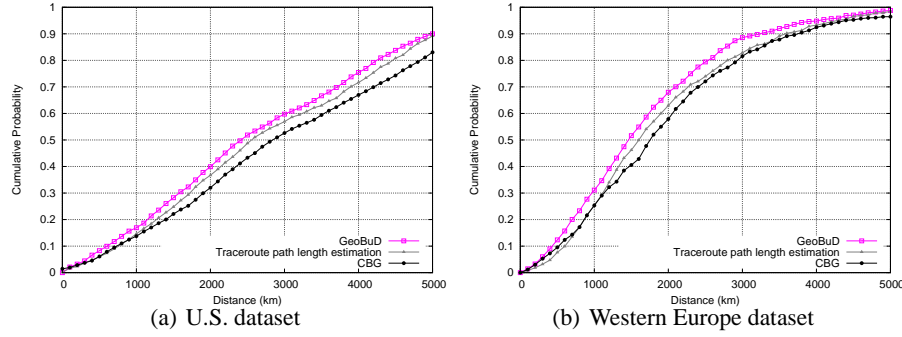
## 5.3 Upper bounds on distance constraints

The fundamental idea of CBG relies on the controlled distance overestimation provided by the distance constraints. The goal is to overestimate in a controlled way so that distance constraints provide the smallest possible region that still encloses the target host. In practice, however, it is important to verify that distance constraints inferred between each landmark and the target hosts effectively provide an upper bound on the actual geographic distance between them. To evaluate whether CBG provides upper bounds on the actual distance, Figure 5 provides the cumulative distribution of the distances for each landmark-target pair.

Figure 5 compares the cumulative distribution of the estimated distances for each landmark-target pair using CBG, the estimation of the traceroute path length, and GeoBuD. The traceroute path length was computed by adding the geographic distances between



**Fig. 4.** Location estimation error for GeoBuD and CBG.



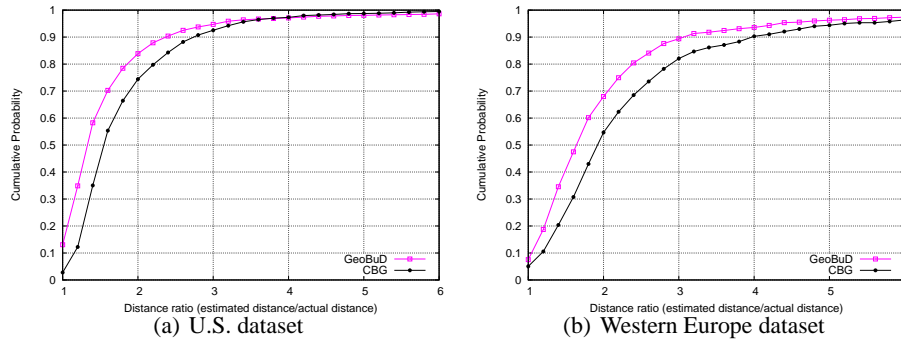
**Fig. 5.** Comparison of distance constraints and path length.

the intermediate nodes along the traceroute which we were able to geolocate. For estimated distances larger than 1000 km, Figure 5 shows that CBG indeed provides an upper bound on the actual geographic distance, as even the estimated geographic length of the traceroute path is smaller. For distances smaller than 1000 km, CBG sometimes is close or below the estimated length of the traceroute path, as the estimated traceroute path length is also an upper bound on the actual geographic distance. Concerning GeoBuD, we observe in Figure 5 that it is a stricter upper bound on the distance than CBG or the estimated traceroute path length.

To understand why GeoBuD outperforms the original CBG, we need to recall how CBG transforms the delay into a distance constraint. For a given landmark, CBG overestimates the actual geographic distance by calibrating its transformation of the delay into a distance constraint by defining the value of  $b$  (see Equation (1)) based on the targets having the lowest delay measurement. This approach has the advantage of providing a conservative upper bound on the distance, as showed in Figure 5. However, its drawback compared to GeoBuD is of obtaining larger confidence regions.

In Figure 6 we plot the cumulative probability of the distance ratio over all landmark-target pairs, *i.e.* the ratio of the estimated distance (with CBG and GeoBuD) to the actual geographic distance. The purpose of Figure 6 is to study how each approach overesti-

mates the actual distance. For instance, if we were to know that all estimated distances have a distance ratio larger than some value, then we could re-calibrate the estimated distance by dividing all distance estimates by this factor. Unfortunately, we can see on Figure 6 that there is a small fraction of the estimated distances that do not overestimate the actual distance (ratio = 1). In fact, 3% of the landmark-target pairs do not overestimate the actual distance for CBG, and 13% for GeoBuD (see Figure 6(a)). In Figure 6(b) we have 5% and 7% for CBG and GeoBuD respectively. If we were to perform this re-calibration for the hosts having a distance ratio of 1, these distance constraints would underestimate the actual distance, potentially leading to an empty confidence region. From Figure 6, we can also see that for a distance ratio smaller than 4, GeoBuD provides a tighter overestimation of the distance than CBG. This is another illustration of the improvement of GeoBuD compared to the CBG approach.



**Fig. 6.** Cumulative probability of distance ratios.

## 6 Conclusion

In this paper we have shown that estimating the buffering delays at intermediate hops along the traceroute between a landmark and a target host enables to improve the accuracy of the geolocation of Internet hosts. Based on traceroute measurements, we estimated the buffering delays at intermediate hops. By combining these buffering delay estimation with a multilateration technique (CBG [10]), we were able to shrink the confidence region where the target host is located. Results show that, with GeoBuD we obtain more accurate location estimation as well.

As further work, we see the implementation of our approach as an on-line tool like GeoLIM [29]. We also aim at converging towards a confidence region as small as possible to provide better location estimation. We might also refine our estimation of the buffering delay by considering the potential existence of a bottleneck link on the path.

## References

1. GeoURL, *GeoURL*, <http://www.geourl.org/>.
2. Net World Map, *Net World Map*, <http://www.networldmap.com/>.

3. Geobytes, Inc., *GeoNetMap*, <http://www.geobytes.com/>.
4. Qwerks, Inc., *WhereIsIP*, <http://www.jufsoft.com/whereisip/>.
5. RegSoft.com, Inc., *ActiveTarget*, <http://www.activetarget.com/>.
6. MaxMind LLC, *GeoIP*, <http://www.maxmind.com/geoip/>.
7. Quova Inc., *GeoPoint*, <http://www.quova.com/>.
8. V. 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.
9. A. Ziviani, S. Fdida, José Ferreira de Rezende, and O. Duarte, "Improving the accuracy of measurement-based geographic location of Internet hosts," *Computer Networks, Elsevier Science*, vol. 47, no. 4, pp. 503–523, Mar. 2005.
10. B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-based geolocation of internet hosts," *To appear in IEEE/ACM Transactions on Networking*, 2006.
11. H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin, "The impact of routing policy on internet paths," in *Proc. of the IEEE INFOCOM'2001*, 2001, pp. 736–742.
12. L. Subramanian, V. Padmanabhan, and R. Katz, "Geographic properties of Internet routing," in *Proc. of USENIX 2002*, Monterey, CA, USA, June 2002.
13. H. Zheng, E. K. Lua, M. Pias, and T. Griffin, "Internet Routing Policies and Round-Trip-Times," in *Proc. of PAM'2005*, Boston, MA, USA, Apr. 2005.
14. 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.
15. University of Illinois at Urbana-Champaign, *IP Address to Latitude/Longitude*, <http://cello.cs.uiuc.edu/cgi-bin/slamm/ip2ll/>.
16. D. Moore, R. Periakaruppan, J. Donohoe, and kc. Claffy, "Where in the world is net-geo.caida.org?," in *Proc. of the INET'2000*, Yokohama, Japan, July 2000.
17. M. Freedman, M. Vutukuru, N. Feamster, and H. Balakrishnan, "Geographic Locality of IP Prefixes," in *Proc. of the ACM Sigcomm IMC'2005*, San Francisco, CA, USA, Oct. 2005.
18. Visualware Inc., *VisualRoute*, <http://www.visualware.com/visualroute/>.
19. CAIDA, *GTrace*, <http://www.caida.org/tools/visualization/gtrace/>.
20. Sarangworld Traceroute Project, <http://www.sarangworld.com/TRACEROUTE/>.
21. P. Bahl and V. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. of the IEEE INFOCOM'2000*, Tel-Aviv, Israel, Mar. 2000.
22. C. J. Bovy, H. T. Mertodimedjo, G. Hooghiemstra, H. Uijterwaal, and Piet van Mieghem, "Analysis of end-to-end delay measurements in Internet," in *Proc. of PAM'2002*, Fort Collins, CO, USA, Mar. 2002.
23. A. Ziviani, S. Fdida, J. Ferreira de Rezende, and O. Duarte, "Toward a measurement-based geographic location service," in *Proc. of PAM'2004*, Apr. 2004.
24. S. van Langen, X. Zhou, and P. van Mieghem, "On the estimation of Internet distances using landmarks," in *Proc. of NEW2AN'04*, St. Petersburg, Russia, Feb. 2004.
25. V. Jacobson, Traceroute software, 1999, <ftp://ftp.ee.lbl.gov/traceroute.tar.Z>.
26. *PlanetLab*, <http://www.planet-lab.org>.
27. NLANR Active Measurement Project, <http://watt.nlanr.net/>.
28. RIPE Test Traffic Measurements, <http://www.ripe.net/ttm/>.
29. GeoLIM Project, <http://planetlab-01.ipv6.lip6.fr:10000/cbg.php/>.
30. N. Spring, D. Wetherall, and T. Anderson, "Scriptroute: A facility for distributed internet measurement flash," in *Proc. of USENIX USITS'03*, Mar. 2003.