

Estimation of Internet Node Location by Latency Measurements – The Underestimation Problem

Dan Komosny

*Brno University of Technology, Brno, Czech Republic
E-mail: komosny@feec.vutbr.cz*

Miroslav Voznak

*VSB – Technical University of Ostrava, Ostrava, Czech Republic
E-mail: miroslav.voznak@vsb.cz*

Kathiravelu Ganeshan, Hira Sathu

*Unitec Institute of Technology, Auckland, New Zealand
E-mail: hsathu,kganeshan@unitec.ac.nz*

Abstract In this paper we deal with discovering a geographic location of a node in the Internet. Knowledge of location is a fundamental element for many location based applications and web services. We focus on location finding without any assistance of the node being located – client-independent estimation. We estimate a location using communication latency measurements between nodes in the Internet. The latency measured is converted into a geographic distance which is used to derive a location by the multilateration (triangulation) principle. We analyse the latency-to-distance conversion with a consideration of location underestimation which is a product of multilateration failure. We demonstrate that location underestimations do not appear in experimental conditions. However with a real-world scenario, a number of devices cannot be located due to underestimations. Finally, we propose a modification to reduce the number of underestimations in real-world scenarios.

Key words: location; service; application; latency; multilateration; underestimation.

1. Introduction and background

Geographic location of Internet nodes is used in many Internet applications, such as on-line credit card fraud detection and prevention, web and other services content personalization, spam filtering, password sharing detection, digital rights management, and any abuse of on-line services. With the Internet of Things that connects small computing devices to the Internet, such as variety of sensor nodes, the knowledge of location of Internet devices becomes even more fundamental.

Native or non-native IP resources can be used for client-independent IP Geolocation, i. e. without any assistance of the node being located. Native are based on common features available in IP systems [20] whereas non-native methods use other resources, such as GPS or radio signal analysis. Non-native approaches are limited in the IP environment¹.

¹For example, the client-dependent GPS system can be used for IP Geolocation. However, there is a line-of-sight path requirement from the node being located to four or more satellites. This is not true for the majority of the IP devices.

Native approaches are known as passive or active (measurement-based). Passive approaches locate a target (IP node being located) by using various databases which store location data, such as WHOIS or DNSLOC [6]. Other databases map blocks of consecutive IP address spaces to geographic locations. However, databases face a number of limitations such as up-to-datedness problem of the stored location data due to new IP addresses assignments and relocations.

The active IP Geolocation involves measurement of communication latency. The latency is measured from a set of landmarks with known geographic positions to the target with unknown location². Simple methods map the target's location to the closest landmark in terms of the lowest latency measured, such as ShortestPing [12]. Similarly GeoPing [21] uses the latency values to build the latency-location profiles. It estimates a location of a target by comparing the latency-location profiles. The target's location is

²The other way is also possible for client-dependent location: a target can estimate its own location by measuring the latency to the landmarks.

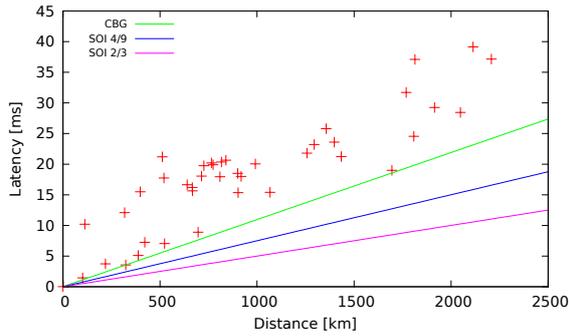


Figure 1. SOI and CBG lines for landmark.

mapped to the location of the landmark with the closest location profile.

The latency can be transformed to a geographic distance [16, 18] which is the base for constraint-based Geolocation. It uses the multilateration principle to estimate a location area of a target [19]. When a unique position is required (for example latitude and longitude), the centroid of the specified area is used at the location of the target. Static and dynamic conversions are used for latency-to-distance transformation. The first one is based on the propagation time of digital information. The first assumptions considered that digital information travels in optical cables at $2/3$ of the speed of light in a vacuum [22]. However, the latency is also a product of other factors, such as processing delays or routing policies [3]. Research into this area has led to a definition of a tighter constraint which is $4/9$ of the speed of light in a vacuum. The Speed of Internet (SOI) method [12] uses this constant. In this paper, we use ‘SOI $4/9$ ’ and ‘SOI $2/3$ ’ notations for the used fractions of the speed of light in a vacuum.

Dynamic conversion uses calibration. The landmark constructs a list of the measured values as shown in figure 1. This figure shows an example the CBG (Constraint-Based Geolocation) [10]. CBG constructs a line which lies under all the points (latency values from landmarks measured) and touches the closest point at the same time. This line is then used to derive the geographic distances for the latencies measured.

The geographic distances obtained from the described methods are then used as input for the multilateration principle to estimate the target’s location. Figure 2 shows an example of IP Geolocation using the CBG method³.

³In our implementation of IP Geolocation, the great circles with radius over 3500 Km were not used for location and they are not shown in the figure.

Other methods such as Octant and Spotter [27, 17] use a similar latency-to-distance calibration based on latency probability models [25, 24], network topology structures, population densities, and city geographic boundaries [1, 5, 2, 17, 4, 13].

In this paper we focus on the location underestimation problem of the active client-independent Geolocation. Location underestimation is a product of multilateration failure when the estimated geographic distances from the landmarks do not intersect to derive a target’ location as shown in figure 2. This failure is caused by an inaccurate latency-to-distance conversion. We found that around 15 % of the location attempts ended with an underestimation. We identified an underestimation if at least one of the great circles around the landmarks did not delimit the correct area of the target location i.e. the target location was not in the resulted area or the area was null.

The paper is structured as follows: The following section describes the related work with a focus on location underestimations. The section ‘Analysis of underestimations’ describes our observation that location underestimations happen with dynamic latency-to-distance conversion in a real-world application. We give an example of an underestimation. We also show how underestimations can be fixed. Next we study how the location accuracy is affected by the underestimation fix. Finally, we conclude the paper.

2. Related work

The need for accurate geolocation of IP nodes without client assistance is an important goal in current Internet research. Location accuracy varies a lot depending on the method used. The accuracy of GeoPing, CBG, Learning-Based, and Octant was discussed in the paper [7]. The median of the location error was between 40-160 km. Paper [23] evaluates the accuracy and reliability of the Spotter method. At city level (40 km) location accuracy was around 30 %. The accuracy also depended on the position of the landmarks⁴. Considering the country-level granularity, Spotter achieved 85 % accuracy of the correct locations. The study [4] evaluated ShortestPing, GeoPing, CBG, and Octant. The results varied from 30 to 200 km of the median location error.

Location underestimations were analysed in paper [12]. The authors found that CBG failed to locate 27 of 128 nodes. Their solution was to use ‘SOI $2/3$ ’. However, this method gives a lower location accuracy. Paper [11] analysed the underestimation problem of the general landmark-based estimation (LBE).

⁴The authors found significant differences for the landmarks in the USA and in Europe.

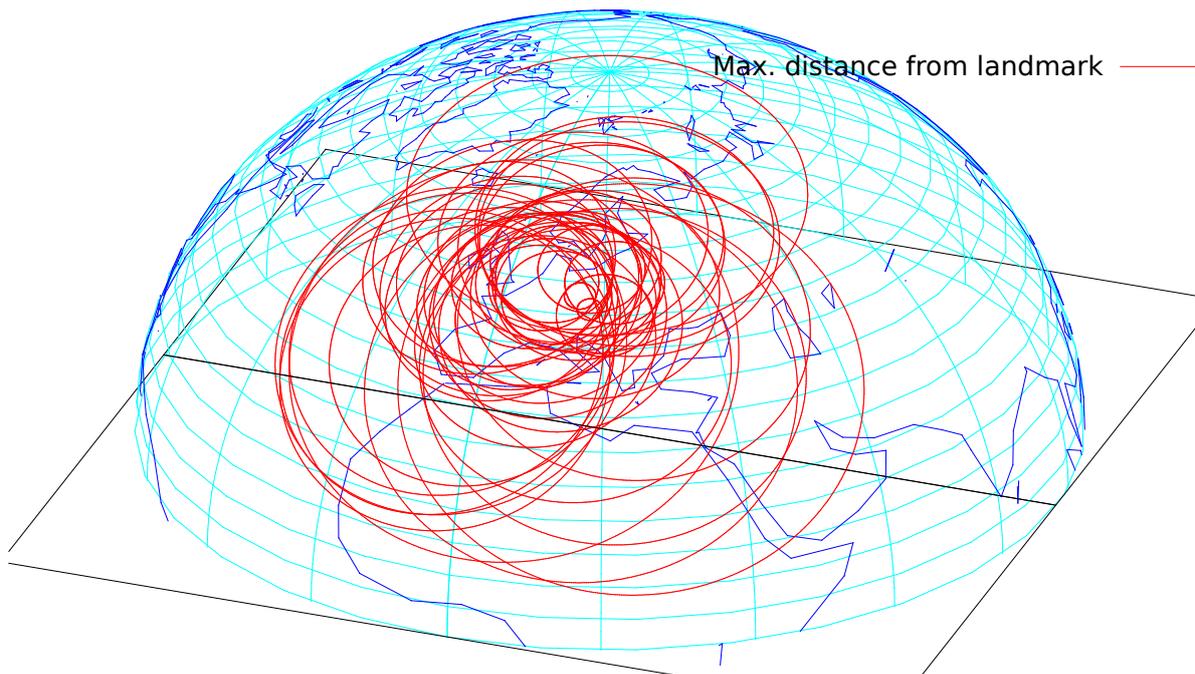


Figure 2. Example location using CBG.

One hundred targets were underestimated due to a wrong calibration using 1411 node pairs. The authors' conclusion and future work was that CBG must be improved to better handle underestimations. They proposed using a different latency-to-distance calibration. Other papers mention the underestimations problem, but they do not present any figures and specific data, such as [10, 9, 8].

3. Analysis of underestimations

In our work we focused on the client-independent geolocation method CBG. CBG is currently recognized as a part of the state-of-the-art IP Geolocation⁵. According to the literature reviewed, the current best active Geolocation approach is described in [26]. It tries to reach street-level location accuracy by employing a three-tier methodology (i) active measurement from geographic landmarks (ii) passive measurement to web servers and (iii) closest node selection. CBG is used in the first-tier to geolocate a target into a specific area.

Our first attempt was to verify the reported accuracy of CBG. For this purpose, we implemented an real-world IP Geolocation system based on the

planetary-scale experimental network PlanetLab. PlanetLab is commonly used as the global geographically-distributed testbed [14]. The developed system works with the PlanetLab nodes in Europe. There are over 300 PlanetLab nodes at over 150 sites⁶. We deployed our software developed for this purpose to each of the nodes. Based on our previous research [15], a number of the nodes were not available for the remote access (ssh) which we needed for latency measurement and software deployment. As a result of this, we used 215 nodes in our experiment.

Our research methodology was the following: Firstly, we did not use the PlanetLab nodes as both the landmarks and targets. The reason was not to have the location results negatively influenced by using both the node types from the same network. The location results when using both the landmarks and the targets from the same networks are of significantly better values. Instead, we used targets outside of PlanetLab. For this, we collected 122 nodes acting as targets. We used sources such as the DNSLOC service [6], and NTP (Network Time Protocol) servers. In this way, we created a real-world scenario for IP Geolocation experiments.

After location of the targets, we noticed that CBG did not work as intended. In a number of cases, about

⁵An on-line CBG Geolocation system can be found here – www.wanmon.slac.stanford.edu/cgi-wrapreflex.cgi

⁶<http://www.planet-lab.eu/>

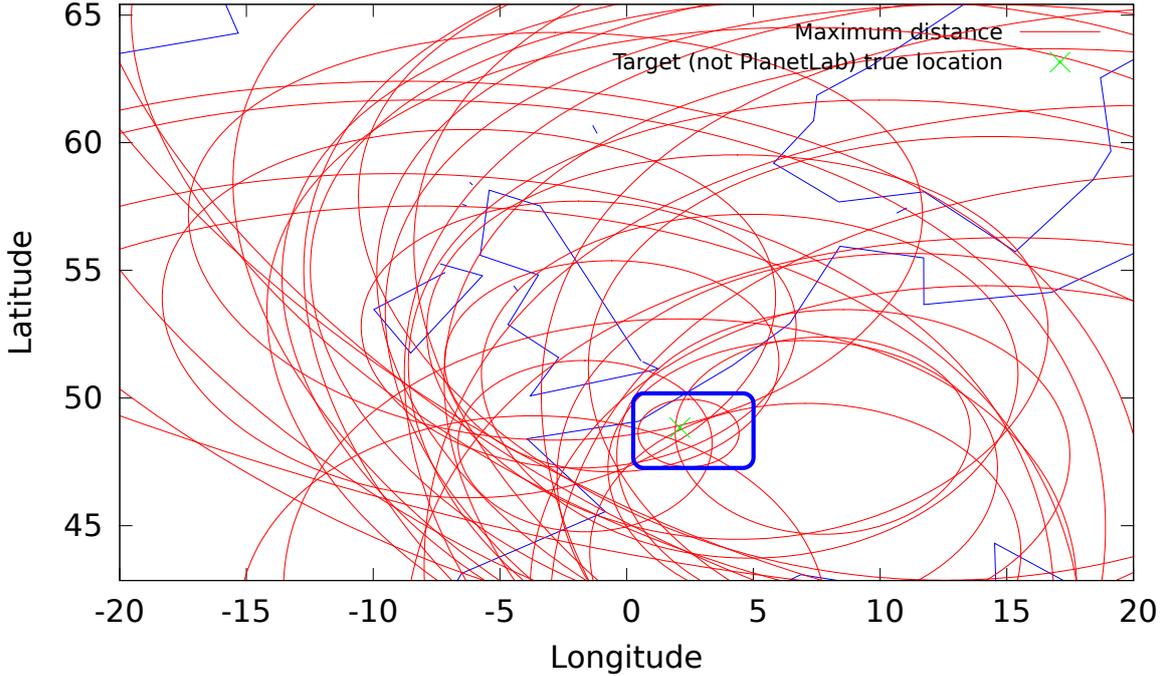


Figure 3. Example of location underestimation using CBG line, TG subset B, target location not estimated.

15 %, we faced a problem of location underestimations. We identified an underestimation if at least one of the great circles around the landmarks did not delimit the correct area of the target location. An example of such underestimation is shown in figure 3 for a target in Warsaw, Poland. We analysed this in more detail and tried to prove a hypothesis that if both the landmarks and targets had been from the same network this issue would not have appeared. To prove this we created two sets of the targets. The first one, subset A, involved targets belonging to PlanetLab. The second one, subset B, involved nodes not belonging to PlanetLab. Moreover, as our intention was to follow the real-world scenario, we also checked all the positions of the nodes. We left only one PlanetLab node at a single location. At the same time, we left only one non-PlanetLab node at a single location. The locations of the PlanetLab nodes and non-PlanetLab nodes were distinct. We considered the minimum distance 25 km between all the nodes. By careful selection we got distinct sets of targets and landmarks. This reduced the number of the nodes, but on the other hand, we created as close to real-world scenario as possible since the location results were not influenced by the similar accuracies of the nodes geographically close to each other. We note that without this selection we obtained about a 20 km better accuracy (on average) than using this modified dataset. An overview of

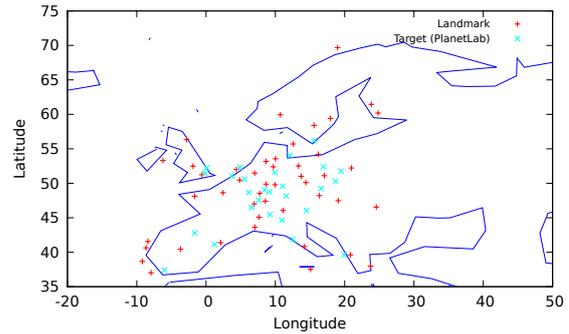


Figure 4. Landmarks and targets, subset A.

the selected nodes (LM – landmark, TG – target) is shown in table 1. The minimum distances achieved are shown in table 2.

The placement of the nodes is shown in figures 4 and 5. The figures differ by the targets. The first figure shows the targets belonging to PlanetLab. The targets do not share their position with any landmark and there is only one target at a single location. The second figure shows the targets outside PlanetLab. Again, any target is not close to any landmark and there is only one target at one location.

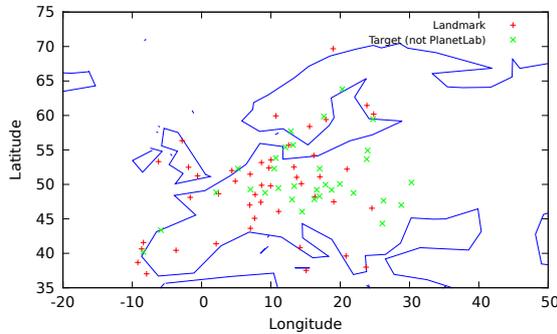
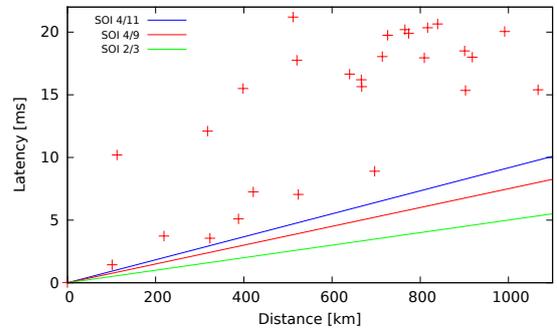
Using the selected nodes we proved that our hypothesis was correct. With subset A, we located all the nodes without any underestimation. On the other

Table 1. Overview of nodes.

Node type	Node count
LM	46
TG all	144
TG set – not same loc. with LM and other TG	59
TG PlanetLab, subset A – not same loc. with LM and other TG	27
TG not PlanetLab, subset B – not same loc. with LM and other	32

Table 2. Minimum distances between nodes.

Between nodes	Minimum distance [km]
LM (PlanetLab)	60
TG (PlanetLab)	35
TG (not PlanetLab)	45
LM (PlanetLab) and TG (PlanetLab)	26
LM (PlanetLab) and TG (not PlanetLab)	25

**Figure 5.** Landmarks and targets, subset B.**Figure 6.** New static calibration line ‘SOI 4/11’ for landmark.

hand, locating targets from subset B resulted in a number of underestimations.

We observed that using a static latency-to-distance conversion reduced the location underestimations in our developed real-world scenario. Both the known static methods, ‘SOI 4/9’ and ‘SOI 2/3’, worked without any underestimation. Moreover, by experiments, we found a new static value of 4/11 of the speed of light which met all our tested criteria. The conversion line ‘SOI 4/11’ is shown in figure 6.

Figure 7 shows a fixed location of the same target which previously failed to be located (Warsaw, Poland, figure 3). The figure also shows the estimated location of the target.

Our aim was to limit the number of location underestimations in real-world Internet applications. We fulfilled this by using the ‘SOI 4/11’ method. Next we compare this method to the original methods CBG, ‘SOI 2/3’, and ‘SOI 4/9’ in terms of location accuracy changes. As the real-world configuration, we used the landmarks from PlanetLab and the targets

outside PlanetLab. We again limited the maximum estimated distance (great circle radius) to 3500 km. The detailed location accuracy results are shown in table 3. The table shows that all the methods give similar results. The maximum median difference is 8 km and the maximum average difference is 19 km. However, a first-look assumption is that ‘SOI 4/11’ should give worse results as a penalty for the location underestimation reduction. The ‘SOI 4/11’ line is below the original CBG line and, thus, the latency-to-distance conversion results in greater maximum distances for the latencies measured. This also holds for ‘SOI 2/3’ and ‘SOI 4/9’. We investigated it and found out that this positive result was caused by the targets additionally located using the ‘SOI 4/11’ method (and other SOI methods) compared to the CBG method. The additionally located targets were found with a smaller location error than the average and they improved the final location accuracy.

The graphical result of the analysis is shown in

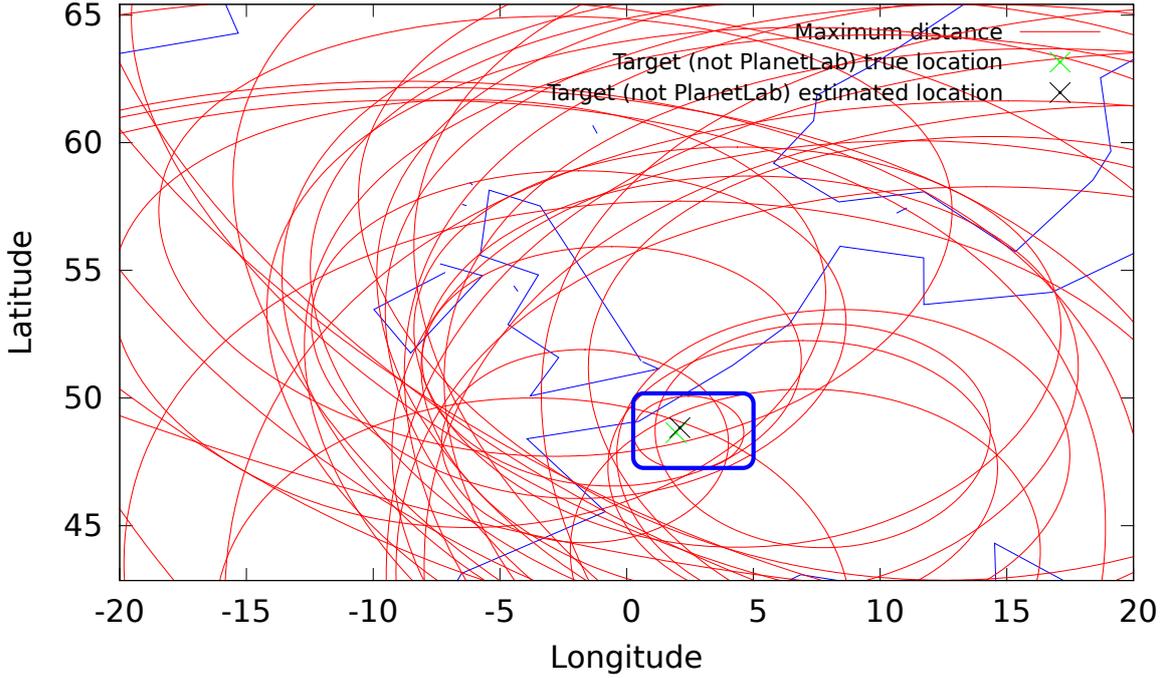


Figure 7. Example of fixed location using line ‘SOI 4/11’, TG subset B, estimated target location shown.

Table 3. Location Accuracy results.

Method	1 st quartile [km]	2 nd quartile [km]	3 rd [km]	Mean	Std. Dev.
SOI 2/3	8	161	226	147	127
SOI 4/9	8	161	226	147	126
SOI 4/11	8	168	226	166	219
CBG	26	160	223	166	207

figure 8. The figure plots a cumulative distribution function of the location accuracy for all the targets used. The function shows the probability of the location error for each method. It can be seen that all the SOI methods, including ‘SOI 4/11’, outperform or equal the CBG method.

4. Conclusion

In this paper, we dealt with the underestimation problem when estimating an Internet node’s position using latency measurements.

We implemented an IP Geolocation system based on the planetary-scale research network PlanetLab. We used 215 PlanetLab nodes as the landmarks and targets, and 122 nodes outside the PlanetLab network as the targets.

We evaluated several scenarios. We located all the targets available, and then we divided the targets into disjunctive sets. We also left the targets which

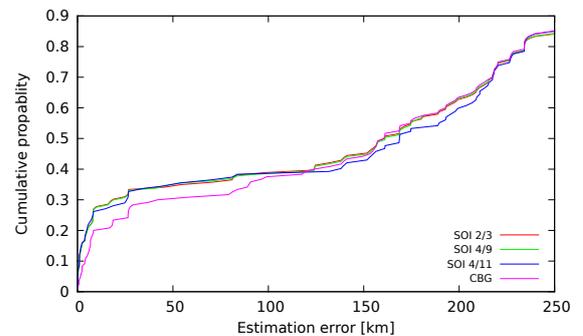


Figure 8. ‘SOI 2/3’, ‘SOI 4/9’, ‘SOI 4/11’, and CBG location accuracy, TG outside PlanetLab.

shared the same geographically-close location with another target or landmark. In this way, we observed that location underestimations did not occur when using both landmarks and targets from PlanetLab. We

explain this by the fact that the high-capacity research PlanetLab network gives better latency measurement results than regular commercial networks. On the other hand when we located the targets outside the PlanetLab network, we observed a number of location underestimations. However, this is the global IP Geolocation scenario.

In order to reduce the location underestimations, we used a static latency-to-distance conversion. We set the slope of the static line as 4/11 of the speed of light in a vacuum. The number of underestimations was reduced to zero and IP Geolocation accuracy was better than with the CBG method which produced the undesired location underestimations.

Acknowledgement: This research was funded by projects SIX CZ.1.05/2.1.00/03.0072.

Bibliography

- [1] M. J. Arif, S. Karunasekera, and S. Kulkarni. GeoWeight: internet host geolocation based on a probability model for latency measurements. In *Proceedings of the Thirty-Third Australasian Conference on Computer Science - Volume 102, ACSC '10*, pages 89–98, Darlinghurst, Australia, Australia, 2010. Australian Computer Society, Inc.
- [2] M. J. Arif, S. Karunasekera, S. Kulkarni, A. Gunatilaka, and B. Ristic. Internet Host Geolocation Using Maximum Likelihood Estimation Technique. In *Proc. 24th IEEE Int Advanced Information Networking and Applications (AINA) Conf*, pages 422–429, 2010.
- [3] C. Bovy, H. Mertodimedjo, G. Hooghiemstra, H. Uijterwaal, and P. Mieghem. Analysis of end-to-end delay measurements in Internet. In *Passive and Active Measurement Workshop-PAM 2002*, 2002.
- [4] B. Eriksson, P. Barford, B. Maggs, and R. Nowak. Posit: An Adaptive Framework for Lightweight IP Geolocation. BU/CS Technical Report, 2011.
- [5] Brian Eriksson, Paul Barford, Joel Sommers, and Robert Nowak. A learning-based approach for IP geolocation. In *Proceedings of the 11th international conference on Passive and active measurement, PAM'10*, pages 171–180, Berlin, Heidelberg, 2010. Springer-Verlag.
- [6] T. Fioreze and G. Heijenk. Extending DNS to support geocasting towards VANETs: A proposal. In *Vehicular Networking Conference (VNC)*, pages 271–277. IEEE, 2010.
- [7] P. Gill, Y. Ganjali, B. Wong, and D. Lie. Dude, Where's That IP? Circumventing Measurement-based IP Geolocation. In *Proceedings of the 19th USENIX Security Symposium*, 2010.
- [8] M. Gondree and Z. Peterson. Geolocation of Data in the Cloud. In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*. ACM, 2013.
- [9] B. Gueye, S. Uhlig, A. Ziviani, and S. Fdida. Leveraging Buffering Delay Estimation for Geolocation of Internet Hosts. In *Proceedings of the 5th International IFIP-TC6 Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*. Springer-Verlag, 2006.
- [10] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based geolocation of internet hosts. *IEEE/ACM Trans. Netw.*, 14:1219–1232, December 2006.
- [11] R. Jayant and E. Katz-bassett. Toward Better Geolocation: Improving Internet Distance Estimates Using Route Traces, 2004.
- [12] E. Katz-Bassett, P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards IP geolocation using delay and topology measurements. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, IMC '06*, pages 71–84, New York, NY, USA, 2006. ACM.
- [13] D. Komosny, J. Balej, M. Simek, P. Dolezel, H. Sathu, and R. Shukla. Cable Length Based Geolocalisation. *Przeglad Elektrotechniczny*, 7a:26–32, 2012.
- [14] D. Komosny, J. Pruzinsky, P. Ilko, J. Polasek, and K. Oguzhan. On Geographic Coordinates of PlanetLab Europe. In *37th International Conference on Telecommunications and Signal Processing (TSP)*, 2014.
- [15] D. Komosny, M. Simek, and G. Kathiravelu. Can Vivaldi Help in IP Geolocation? *Przeglad Elektrotechniczny*, 5:100–106, 2013.
- [16] O. Krajsa and L. Fojtova. RTT measurement and its dependence on the real geographical distance. In *Proceedings of 34th International Conference on Telecommunications and Signal Processing*. Brno University of Technology, 2011.
- [17] S. Laki, P. Matray, P. Haga, T. Sebok, I. Csabai, and G. Vattay. Spotter: A model based active geolocation service. In *INFOCOM*, pages 3173–3181. IEEE, 2011.
- [18] P. Moravek, D. Komosny, R. Burget, J. Sveda, T. Handl, and L. Jarosova. Study and performance of localization methods in IP based networks: Vivaldi algorithm. *Journal of Network and Computer Applications*, 34(1):351–367, 2011.
- [19] P. Moravek, D. Komosny, and M. Simek. Multilateration and Flip Ambiguity Mitigation in Ad-hoc Networks. *Przeglad Elektrotechniczny*, page 9, 2012.
- [20] P. Moravek, D. Komosny, M. Simek, J. Sveda, and T. Handl. Vivaldi and other localization methods. In *Proceedings of the International Conference on Telecommunications and Signal Processing (TSP)*, Dunakiliti, Hungary, 2009. Asszisztencia Szervezo Kft. Budapest.
- [21] V. Padmanabhan and L. Subramanian. An investigation of geographic mapping techniques for internet hosts. *SIGCOMM Comput. Commun. Rev.*, 31:173–185, August 2001.

- [22] R. Percacci and A. Vespignani. Scale-free behavior of the Internet global performance. *The European Physical Journal B - Condensed Matter and Complex Systems*, 32:411–414, 2003.
- [23] Y. Shavitt and N. Zilberman. A Geolocation Databases Study. *IEEE Journal on Selected Areas in Communications*, 29(10):2044–2056, 2011.
- [24] M. Voznak, M. Halas, F. Rezac, and L. Kapicak. Delay variation model for RTP flows in network with priority queueing. In *10th WSEAS International Conference on EHAC'11 and ISPRA'11*. WSEAS, 2011.
- [25] M. Voznak, R. Jan, and H.A. Ilgin. Network delay variation model consisting of sources with Poisson's probability distribution. In *Proceedings of the 15th WSEAS International Conference on Computers*, 2011.
- [26] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang. Towards street-level client-independent IP geolocation. In *Proceedings of the 8th USENIX conference on Networked systems design and implementation*, NSDI'11, pages 27–27, Berkeley, CA, USA, 2011. USENIX Association.
- [27] B. Wong, I. Stoyanov, and E. Sirer. Octant: a comprehensive framework for the geolocalization of internet hosts. In *Proceedings of the 4th USENIX conference on Networked systems design & implementation*, NSDI'07, pages 23–23. USENIX Association, 2007.