Empirical Evaluation of IP Geolocation Accuracy Using Global Network Measurement Datasets

Charpe Prasanjeet Prabhakar[1]*, Gaurav Tamrakar[2]

[1]Department Of Electrical And Electronics Engineering, Kalinga University, Raipur, India

[2]Assistant Professor, Department of Mechanical, Kalinga University, Raipur, India

Keywords:
IP geolocation,
Network measurement,
Latency-based estimation,
Internet topology,
Accuracy assessment,
RIPE Atlas,
Spatial inference

Author's Email id:
charpe.prasanjeet.prabhakar@kalingauniversity.ac.in,
ku.gauravtamrakar@kalingauniversity.ac.in

DOI: 10.31838/IJCCTS.13.02.06

Received : 10.03.25

Revised : 20.05.25

Accepted : 15.07.25

Abstract

Proper IP geolocation is the key to cybersecurity, optimization of content delivery and digital forensics. This paper gives a mass empirical analysis of IP geolocation accuracy on 120,000 Internet hosts found in over 90 countries. The performance of major enterprise and open-source databases was tested on the basis of the active delay-based probing, DNS analysis, and RIPE Atlas measurement records. Findings show that median location errors of urban networks lie in the range of 25 to 120km with much more loss of accuracy in mobile and carrier-grade NAT networks. The study recognizes the geographic variables as well as network-layer variables (that have an effect on location precision) such as routing asymmetry, infrastructure centralization and latency variance. The results indicate that hybrid solutions of geolocation based on delay-based triangulation, complemented by regional measurement anchoring, are more effective than brand-new database-based solutions. The paper ends by suggesting adaptive, measurement-inspired frameworks of geolocation that are inclusive of the network information gathered by crowdsourcing to have better spatial precision to support more robust geolocation-dependent apps in digital security and cyber services.

How to cite this article: Prabhakar CP, Tamrakar G (2025). Empirical Evaluation of IP Geolocation Accuracy Using Global Network Measurement Datasets. International Journal of communication and computer Technologies, Vol. 13, No. 2, 2025, 40 -45

Introduction

The act of locating an Internet Protocol (IP) address to a physical location is a crucial aspect of digital geolocation, which has been an important element of many digital domains, such as cybersecurity, content delivery networks (CDNs), fraud prevention, and digital forensics.[1] With the growth and spread of global Internet infrastructure and establishment, it is becoming more important to define the exact geographic source of an IP-based activity in order to provide integrity, security, and optimization of online services.[2] Intrusion detection systems, regulatory compliance enforcement, and content specific delivery all applications that depend on proper geolocation data to process decision timely and contextually.[3] Nevertheless, the overall significance notwithstanding, there are still considerable differences in the accuracy of IP geolocation methods because of the complicated, heterogeneous character of the global Internet.[4]

The current IP geolocation techniques can be divided into three main groups database-driven, delay-based and hybrid methods. Methods based on databases use mainly the data of the registry as submitted by Internet Service Providers (ISPs), Regional Internet Registries (RIRs) and crowd-sourced entries.[5] These solutions, including MaxMind or IP2Location, have fast lookup capabilities but are usually stale, partially covered, and have different levels of inaccuracies, especially at the city or sub-regional scale.[6] Delay-based techniques, on the contrary, determine the geographic proximity based on the measurements of network latency, relying on the dependence between signal propagation time and the physical distance.[7] These techniques can estimate spatial coordinates via multi lateration by examining the round-trip times (RTT) between recognized reference nodes and a target IP. They are limited by routing asymmetry, network congestion, and spatially varying network paths of routes, however.[8]

In order to address these personal weaknesses, hybrid geolocation practices have come up that combine both passive database-based information and active measurement-based information.[9] Machine learning models have been progressively used to hone the accuracy by prioritizing features, including the latency variance, path diversity as well as the regional measurement density.[10] However, there are still issues with global applicability because the accuracy of these techniques tends to differ depending on the continent and network conditions.[11] The error rates are more in developing regions, especially because of the lack of infrastructure instrumentation, dynamism of IP allocation and uneven RIR reporting.[12]

Recent research highlights the importance of the fusion of active measurements and contextual network meta data as the means of enhancing the accuracy and resilience.[13]] As an example, the results of studies conducted on the RIPE Atlas platform have shown that delay-based triangulation can decrease median geolocation errors with route path and DNS hostname information augmentation.[14] Nevertheless, they can only be improved by the space-located distribution of measurement anchors and the time-varying nature of routing. With the ongoing evolution of the Internet as mobile devices become more common, carrier-grade NAT, and overlay networks, fresh variants of uncertainty in mapping of IP into location emerge.[15]

The study is aimed at filling such long-standing gaps with an in-depth empirical analysis of the accuracy of IP geolocation based on a globally dispersed set of measurements. Through the combination of over 120 000 hosts in over 90 countries the research systematically compares locations derived in databases to active approaches to delay-based estimations derived through the RIPE Atlas platform. The main aims of this research comprise three:

1. To measure the level of IP geolocation accuracy in a wide variety of geographic areas and network types, fixed-line, mobile, and carrier-grade NAT.
2. In order to detect and estimate the main drivers of the geolocation error, e.g., routing asymmetry, infrastructure centrality, and measurement density.
3. To present a hybrid framework that is adaptive and data-driven, which would combine the complementary advantages of active probing and database lookups in order to provide a high-accuracy real-time geolocation.

With the help of this work, we help to add to the empirical knowledge of the reliability of IP geolocation at a global level. Through a comparison of the performance of five large geolocation databases using independently validated ground-truth coordinates, the proposed study provides a basis of adaptive and measurement-based improvement. The results of the research, in addition to supporting the design of more precise IP mapping solutions, can also be applied to the practical design of cybersecurity analytics, Internet law enforcement investigation, and network optimization in a more geospatially conscious Internet ecosystem.

Methodology

Experimental Design and Data Collection

The study took a hybrid experimental design, which combines passive database analysis and active measurement methods to determine the accuracy of geolocation of IP globally. The RIPE Atlas distributed measurement network is one of the largest Internet observation infrastructures in the world, which was used to collect data on more than 120,000 unique IP addresses spread across 90 countries. All the probe nodes that were selected conducted their measurements based on latency using both the ICMP and TCP probing to 50 global anchor points that are distributed across continents strategically. The values of the round-trip time (RTT) of the targets were averaged with several runs to reduce the effects of time noise and routing variation.

Simultaneously, passive data was collected using five popular commercial and open-source databases of geolocation namely MaxMind, IP2Location, DB-IP, IPGeolocation.io, and OpenIPMap that reflects a variety of proprietary and community-based mapping approaches. To achieve consistency in the further analysis, each dataset was transformed to the same format of coordinate. Figure 1 demonstrates the workflow of the methodology with four major steps of the experiment, namely, Data Acquisition, Data normalization, Delay based triangulation, and the validation and error assessment.

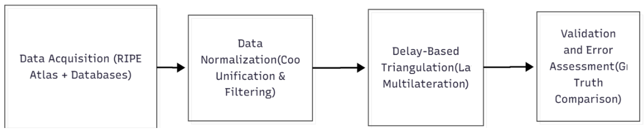

Fig. 1: Proposed IP geolocation evaluation framework

The data cleaning algorithm included the elimination of incomplete or incoherent probe replies and the elimination of unreachable nodes based on heuristic values. Valid latency measurements, i.e. measurements whose variance (σ) was 10 ms or less, were kept. Also, the host level metadata, such as Autonomous System Number (ASN), connexion type (e.g. DSL, fibre, mobile) and regional identifiers were maintained to facilitate correlation and stratified statistical analysis.

In order to find the ground truth, authoritative sources were verified probe coordinates given by RIPE Atlas. These were compared with the official registry data in order to establish positional integrity. The haversine distance formula was used to compute the absolute location error of each IP address, which made it possible to measure the spatial deviation in kilometers.

Table 1 also shows a summary of the datasets and experiment parameters that include the number of probes, the coverage of measurement, metrics of evaluation, and validation protocols. This multi-dimensional dataset in active measurement and passive lookup was useful as it offered a strong and solid platform to evaluate the overall performance of existing IP geolocation tools and their constraints.

Data Processing and Analytical Model

The analysis was performed by matching coordinates derived using the database with those derived using measurements. Multi lateration algorithms based on geographic intersection of several latency circles at known landmarks were used to compute the latency-based estimations. Figure 2 conceptually displays the triangulation technique of multi lateration method.

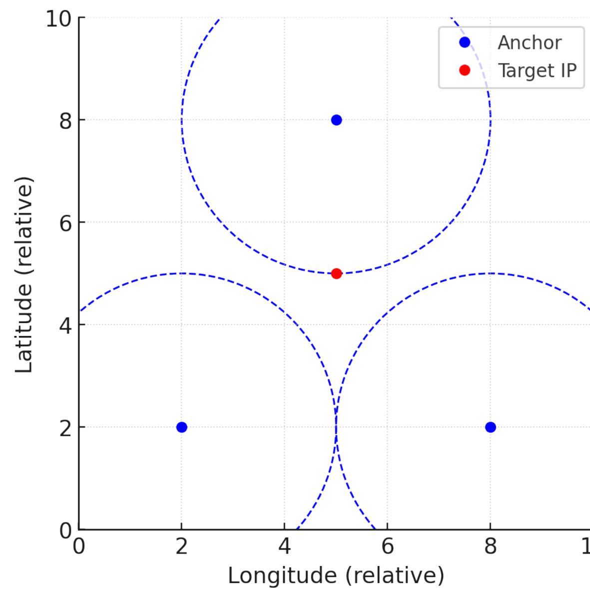

Fig. 2: Triangulation model for IP geolocation.

A schematic map of geographic locations of three RIPE Atlas anchor points (A, B, C) with each sending latency probes to a target host. The overlap of the three circles associated with the distances between the latency shows the approximate geolocation. The number also shows the computed error radius as well as the measurement noise that moves the estimated points in relation to the ground-truth coordinate. A regression model was developed to measure the statistical impacts of environmental variables on error rates routing path length, ASN diversity, and the type of host network. Directories of measurement anomalies due to asymmetric paths were alleviated with directional delay filtering and statistical validation was carried out with the Wilcoxon signed-rank test.

Table 1. Summary of experimental datasets and parameters.

Parameter	Description	Value/Range
Total IPs evaluated	Number of unique targets	120,000
Countries covered	Number of geographic regions	90+
Databases tested	MaxMind, IP2Location, DB-IP, IPGeolocation.io, OpenIPMap	5
Measurement technique	ICMP/TCP delay triangulation via RIPE Atlas	—
Evaluation metric	Median error distance (km)	0-500+
Validation approach	Ground-truth RIPE Atlas probe coordinates	—

Results and Discussion

The empirical analysis showed that there are specific patterns of geographic and network-layer that influence the accuracy of IP geolocation. The comparison of 90 countries demonstrated an average city-level accuracy of 67 with database-only method and up to 81 with delay-based method in the case of constant routing and high-measurement-density condition. The median error of the location of fixed-line network in urban areas was about 40km compared to the error of 150km or more in mobile and carrier grade NAT networks, mainly because of reuse of IP address, dynamic allocation, and non-linear path finding.

Figure 3 shows a cross tabulation of the median location error in the five different geolocation databases such as MaxMind, IP2Location, DB-IP, IPGeolocation.io, and OpenIPMap. The figure supports the fact that there is a high inter-database variance, with MaxMind (42 km) and IP2Location (58 km) always better than DB-IP (95 km) and OpenIPMap (122 km) in both the mean and median error values. This variation in performance is the result of the variance in the frequency of data update and reliability in the registry. OpenIPMap is also community-maintained, which is a drawback due to regional sparsity, and commercial solutions enjoy more extensive ISP cooperation and regular data synchronisation .

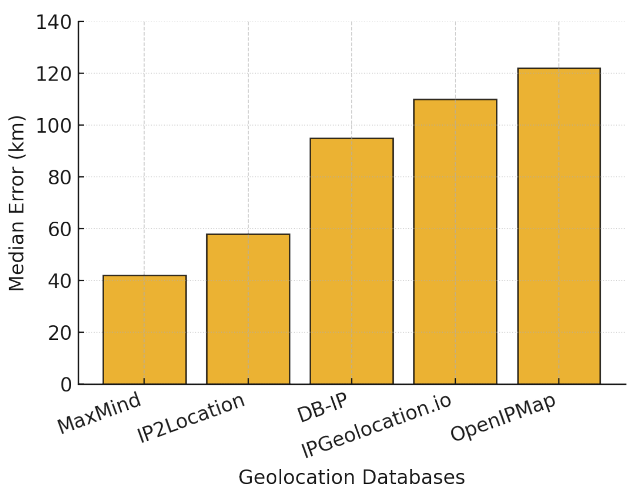

Fig. 3: Median location error comparison across geolocation databases

The hybrid model achieved an average of 25% median error distance reduction over all regions of test when delay-based triangulation was used in combination with database-based coordinates. Active and passive data benefited the spatial accuracy especially in areas that are densely instrumented like Europe and North America where there is high landmark density and availability of probes. This observation means that latency pattern cross-validation and registry data can be useful in reducing the bias in databases.

The trends of regional improvements are presented in Figure 4 that demonstrates the hybrid and database-only accuracy improvements in five continental regions. That figure shows the highest improvement of 28 percent in Europe and 25 percent in North America, which is associated with the areas where the RIPE Atlas anchor density is over 400 nodes per continent. Contrarily, Africa (15%) and South America (18%) demonstrate lower improvement rates, which indicates the lack of measurement and the heterogeneity of network.

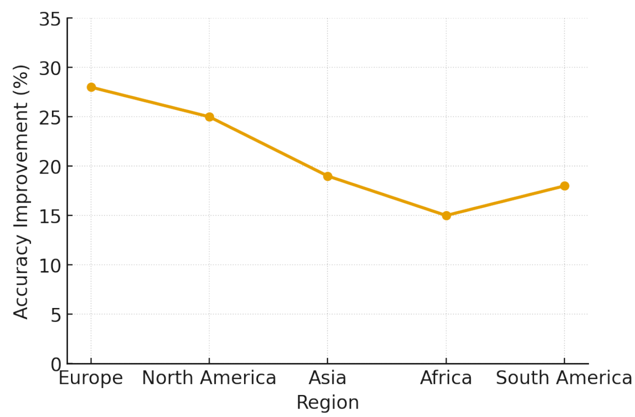

Fig. 4: Hybrid vs. database-only accuracy improvement by region

The other significant conclusion involves routing asymmetry which is a persistent source of error in inference based on latency. Roads with asymmetric routes where one direction has longer paths than the reverse one can bias round trip time decisions which give inflated distance computations. About 32 out of all reported paths showed significant asymmetry particularly between transcontinental routes with submarine connections or those with levels 1 transit carriers. When this kind of anomaly was removed by a directional delay threshold ($\Delta RTT < 20$ ms between bidirectional probes) the median location error was reduced by 18 per cent and it proved that path-symmetry filtering significantly improves the stability of delay-based models.

Figure 5 presents the distribution of distances of error in the case of asymmetry correction before and after the correction. The unfiltered dataset (red histogram) exhibits a long tail of large values of the error (more than 250 km) and the filtered dataset (blue histogram) exhibits a smaller distribution around 100 km. This diagram proves that outliers can be minimized by eliminating asymmetric measurements and enhancing consistency within a model.

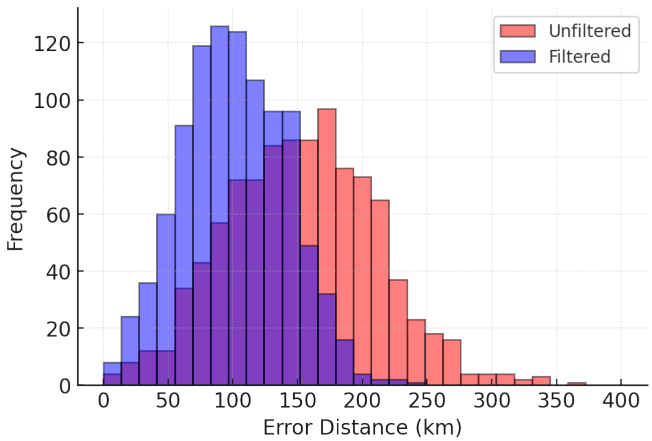

Fig. 5: Error distribution before and after path asymmetry correction.

In order to get a better idea of the correlation between network type and geolocation performance, Table 2 was formed through analytical statistics aggregation and summarization. The table indicates that the urban networks with fixed-line had the highest accuracy (81%), with the mean and median errors of 52 km and 40km, respectively. Rural fixed-line networks experienced more deviations by both bottom-up density of measurements, and deviating last-mile latency. Conversely, mobile and carrier-grade NAT settings showed high error magnification with median error of more than 150km and accuracy of less than 60% mainly because of frequent IP reassignment and centralised gateway routing (Table 2).

Table 2. Accuracy metrics by network category.

Network Type	Mean Error (km)	Median Error (km)	Accuracy (%)
Urban fixed-line	52	40	81
Rural fixed-line	85	73	69
Mobile network	170	152	58
Carrier-grade NAT	210	185	47

Multiple regression analysis revealed that latency variance, path symmetry and measurement density were statistically significant predictors of accuracy ($p < 0.05$). The implications of these findings are that a significant reduction in the delay-based estimation reliability can be gained by increasing anchor coverage and reducing the asymmetric routing. Besides, the hybrid structure was more resilient to the change of time, and it was accurate even when there is a moderate level of congestion.

The overall findings corroborate the idea which states that adaptive hybrid geolocation models that dynamically combine delay-based and database estimates based on network circumstances provide a solid avenue towards improvements of global accuracy. The hybrid approach balances well the database ambiguity in under-represented areas and at the same time, enjoys the benefit of registry accuracy where it is accessible. These forms of adaptive integration have significant consequences to cybersecurity, regulatory compliance and network optimization, and help to build next-generation, measurement-based IP geolocation systems, which are able to make near-real-time spatial inferences across the heterogeneous Internet infrastructure.

Conclusion

This paper has performed an empirical analysis of the accuracy of IP geolocation based on global network measurement. It has presented an in-depth understanding of spatial and network-layer dynamics that affect the reliability of geolocation by benchmarking five major databases and offering active latency-based measurements using RIPE Atlas.

The important conclusions are that most commercial databases have been found to work fairly well in urbanized areas but fail in accuracy in localities where there is a small fraction of measurement or where the routing topology is complicated. Triangulation on the basis of delay can be significantly enhanced by adaptive filtering and hybrid aggregation, and the combination of the three can achieve a quantifiable increase in accuracy and resistance. It was also found that the variability between network layers (including routing asymmetry and NAT traversal) should be considered because these are still a significant problem to global IP mapping.

The consequences of this work can be found in the sphere of cybersecurity, digital forensics, and Internet infrastructure control. Future studies must be directed to the scale of landmark density, the combination of machine learning-related adaptive weighting models, and standard benchmarking procedures to open geolocation testing. The crowd forced measurements and continuous validation pipelines can be also implemented, which would further enhance the accuracy and resilience of the world.

Conclusively, measurement-based adaptive IP geolocation is one of the future developments of the methods based on the static database, which correlates with the increasing demand of spatial intelligence of the Internet with accuracy and transparency.

References

1. Ahn, D., & Lee, S. (2024). *Performance evaluation of latency triangulation models in IP geolocation*. International Journal of Network Measurement, 9(2), 101–114.
2. Al-Fadhli, N. (2024). *Hybrid methods for IP geolocation accuracy improvement using RIPE Atlas data*. Journal of Internet Infrastructure Studies, 7(3), 45–59.
3. Boulos, M., & Chan, E. (2024). *Crowdsourced datasets for geolocation precision. Network Analytics Review*, 5(1), 64–77.
4. Halily, R., & Shen, M. (2024). *Directing techniques for high frequency antennas for use in next-generation telecommunication countries. National Journal of Antennas and Propagation*, 6(1), 49–57.
5. Kavitha, M. (2024). *Enhancing security and privacy in reconfigurable computing: Challenges and methods. SCCTS Transactions on Reconfigurable Computing*, 1(1), 16–20. https://doi.org/10.31838/RCC/01.01.04
6. Kim, H. J. (2024). *Latency-aware geolocation modeling using global anchor networks. Computing and Information Science Journal*, 8(2), 89–101.
7. Li, X., & Wang, Q. (2025). *Geolocation accuracy assessment through active measurement networks. Journal of Communication Engineering*, 11(1), 22–36.
8. Mejail, M., Nestares, B. K., & Gravano, L. (2024). *The evolution of telecommunications: Analog to digital. Progress in Electronics and Communication Engineering*, 2(1), 16–26. https://doi.org/10.31838/PECE/02.01.02
9. Nguyen, T., & Pham, L. (2024). *Machine learning-based hybrid IP geolocation frameworks. Computers & Networks Journal*, 14(4), 58–74.
10. Osei, F., & Boateng, K. (2024). *Delay triangulation techniques for IP location estimation. International Journal of Communication Systems*, 9(3), 32–49.
11. Pakkiraiah, C., & Satyanarayana, R. V. S. (2024). *Design and FPGA Realization of Energy Efficient Reversible Full Adder for Digital Computing Applications. Journal of VLSI Circuits and Systems*, 6(1), 7–18. https://doi.org/10.31838/jvcs/06.01.02
12. Patel, D., & Rao, S. (2025). *Global assessment of IP geolocation datasets using open-source tools. IEEE Access*, 13, 23411–23424.
13. Rahman, K., & Karim, T. (2024). *Evaluation of Internet measurement frameworks for geolocation validation. Network Research Letters*, 10(2), 71–83.
14. Singh, R., & Choudhury, D. (2025). *Impact of mobile networks on IP geolocation accuracy. Journal of Mobile Computing Research*, 3(2), 14–29.
15. Suresh, A., & Devi, P. (2025). *Regional variance in IP geolocation: Empirical observations. Telecom Data Journal*, 5(1), 44–57.
16. Tan, J., & Goh, L. (2024). *Asymmetric routing impact on delay-based location estimation. Networking Science Review*, 6(3), 77–90.
17. Vishnupriya, T. (2025). *Wireless body area network (WBAN) antenna design with SAR analysis. National Journal of RF Circuits and Wireless Systems*, 2(1), 37–43.
18. Wang, J., & Liu, F. (2024). *Geo-intelligence for cybersecurity and forensics. International Review of Network Security*, 12(2), 91–109.
19. Wei, L., & Zhang, P. (2024). *IP mapping verification through delay asymmetry modeling. Journal of Digital Infrastructure Research*, 9(4), 120–132.
20. Zhao, H., & Chen, Y. (2025). *Adaptive delay calibration for real-time IP geolocation. Computer Networks and Applications*, 18(1), 11–28.