DMPedia Lecture Notes in Multidisciplinary Research,
Volume 1, pp. 615-627, 2026

International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

IPv6 LANMARKER-PLUS : ENHANCING GEOLOCATION ACCURACY AND EFFICIENCY VIA STAGED PROBING AND DENSITY-BASED CLUSTERING

S.Rajesh, V.Nithya,C.Rethiga,S.Sujitha

Department Of Information Technology

Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu.

(An Autonomous Institution Affiliated By Anna University, Chennai)

srajesh@mepcoeng.ac.in , nithyavelmurugan_bit27@mepcoeng.ac.in,
ethiga0405_bit27@mepcoeng.ac.in , sujicsk7_bit27@mepcoeng.ac.in

ABSTRACT

The accurate estimation of the street-level geolocation of IPv6 addresses poses a significant challenge due to the vast address space and the rotation of prefixes driven by privacy concerns. Methods like IPv6Landmarker achieve substantial improvement compared to prior IPv4 techniques; however, these rely on brute-force querying and heuristic-based clustering, which are computationally inefficient and vulnerable, respectively. In this paper, we present IPv6Landmarker-Plus-an enhanced extension that significantly improves IPv6 landmark mining in terms of both accuracy and resource efficiency. Our system utilizes Java with a MongoDB architecture to filter EUI-64 addresses out of active IPv6 hitlists and extract associated WAN MACs for cross reference with the Wigle.net database. Our contributions include two major algorithmic enhancements: 1) Staged Offset Probing (SOP): a probabilistic query model that, for high probability offsets such as -1, +5, and 0, reduces the number of external API calls by more than 80%, significantly reducing latency and resource consumption. 2) DBSCAN-Centric Filtering: this replaces heuristic clustering with Density-Based Spatial Clustering to accurately identify the true coordinate cluster while robustly rejecting noise from adjacent network devices, leading to direct improvements in the mean geolocation error. Real-world validation using RIPE Atlas ground truth probes confirms that the DBSCAN-based selection of landmarks achieves a lower mean error and geolocation success rate compared to the original technique in IPv6Landmarker.

Keywords: *Ipv6 , Ipv6 Geolocation-network landmark, WiFi BSSID-landmark mining.*

1. Introduction

IP geolocation is a fundamental capability required by a wide range of security and network applications, such as fraud detection, content customization, and CDNs1. The accuracy of these services relies heavily on the availability of a large pool of high-quality network landmarks-nodes with known, stable geographic location and identifiers2. However, the IPv6 geolocation faces unique and serious challenges that are hard to address with traditional approaches developed for IPv4 333. There are two main issues that stand in the way of creating sufficiently dense and accurate landmark sets for street-level geolocation-mean error is usually below (mean error typically under 10 km: Huge Address Space: The huge size of the IPv6 address space (2^128) makes brute-force scanning-based approaches infeasible to find active endpoints5. Address Instability: The client privacy mechanisms, like IPv6 address prefix rotation, cause the IP addresses of known landmarks to change periodically and render them unreachable and unstable. The solutions developed so far, like GeoCAM7 and LandmarkMiner8, have targeted IPv4 and cannot be scaled up to the IPv6 environment. While the approaches in IPvSeeYou and IPv6Landmarker leverage the unique features of EUI-64 addresses-the MAC address is embedded in the IID-the IPvSeeYou approach can cover only

3% of the registered OUIs limiting the amount of resulting landmarks. Moreover, its efficiency is hampered by a rigid and nonprobabilistic methodology of discovering the corresponding BSSIDs.

2. Related Work

A literature survey serves as both a summary and a critical analysis of research that has previously been undertaken in a specific domain—in this case, IPv6 geolocation. Its main objective is to spotlight transformative advancements, showcase ongoing challenges, and illustrate how emerging methodologies, such as GDD-Geo and related network geolocation algorithms, fit into the evolving landscape of the 2 field. By examining the body of work leading up to the present study, this section aims to situate the current research within broader academic and practical contexts, highlighting the necessity for greater accuracy, stability, and scalability in network geolocation.

2.1 GDD-Geo and State-of-the-Art Methods

Within IPv6 geolocation research, various strategies have emerged to improve the granularity and reliability of IP-based location determination, especially as IPv6 adoption accelerates. Landmark mining remains a foundation—these are methods to identify network nodes whose locations are already known and can be used to anchor subsequent geolocation inferences. Parallel to this, researchers have explored both passive and active approaches; some rely on ongoing traffic flows and latency analysis, while others use targeted probing or hybrid methodologies to triangulate positions. Machine learning and advanced graph-based models increasingly underpin the most promising solutions, pushing accuracy and adaptability to new heights. Among these, the GDD-Geo approach is distinguished by its use of graph dual decomposition, which incorporates not just the immediate network topology but also the spatial relationships between distributed landmarks. This dual focus allows GDD-Geo to achieve high localization accuracy, even when the data on anchoring nodes is sparse—an advantage over earlier algorithms that struggled with limited anchor density or depended heavily on frequently outdated external databases. A comparative evaluation shows that while legacy methods like SLG/MLP-Geo remain reliable in dense anchor scenarios, and MAC-based models such as IPvSeeYou garner fine accuracy in privacy-permissive settings, they are often limited by either error rates in sparse environments or practical deployment constraints. In contrast, GDD-Geo can reduce location estimation errors by up to 60% over previous systems when working with a judicious selection of anchor nodes.

2.2 Addressing Research Gaps and Motivations

Despite prior innovations, key obstacles persist in the quest for universally precise geolocation: low city and street-level precision when landmark nodes are few, an overreliance on external and sometimes static geolocation databases, and the inherent difficulty of adapting to the constantly shifting topology of IPv6 networks. GDD-Geo responds to these concerns with algorithms and frameworks explicitly designed to scale to large, dynamic internet address spaces while maintaining spatial accuracy. Its strategy is to maximize both the spatial distribution and stability of anchor nodes, balancing operational feasibility with technical rigor in real-world, ever evolving deployment environments.

2.3 Stable IPv6 Landmark Representation

Accurate geolocation hinges not just on quantity, but also on the stability of landmark nodes—network elements whose locations can reliably serve as ground truth for algorithmic inference. Traditional landmark selection, often predicated on MAC address mapping or mere proximity, suffered from both instability and poor spatial coverage, resulting in unreliable results, especially as IP addresses change or churn due to network events. Research on stable IPv6 landmark representation has thus shifted focus to multi-feature clustering techniques, as exemplified by the landmark-v6 approach. Here, candidate landmarks are selected based on multiple correlated attributes—including address traits, observed geographical features, network latency, and historical connectivity—to produce highly stable and widely distributed anchors. In empirical studies, this clustering-driven methodology consistently outperforms older, more simplistic selection strategies, offering improved landmark survivability and coverage for large-scale geolocation deployments.

2.4 Hybrid and Advanced Landmark Mining: DualS-Geo

Recent frameworks such as DualS-Geo advance the field further by tapping into dual-stack (IPv4/IPv6) devices, mining for landmarks that appear in both address spaces. This dual protocol strategy effectively multiplies the anchor pool, since reliable IPv4 geolocation data can be leveraged to inform IPv6 mapping. By integrating cross-protocol correlations such as domain name associations and delay measurements—DualS Geo substantially boosts street-level precision and coverage, particularly in areas where IPv6-only anchors are scarce. Yet, this methodology is not without its caveats, notably the potential to inherit errors or outdatedness from legacy IPv4 datasets and the complexity of accurately correlating information across network layers.

2.5 Beyond Graphs: Hypergraph Learning in IPv6 Geolocation

As the address space and topology of the Internet grow ever more complex, classical graph-based models sometimes fall short, especially in capturing non-adjacent, yet geographically correlated, hosts that evade simple connectivity-based inference. Enter HGLGEO, which frames the problem as node regression on a hypergraph—effectively allowing the algorithm to group and learn from sets of hosts with shared spatial attributes, even if they are not topologically adjacent. With hypergraph convolution, this model enables nodes that are geographically close, but not directly networked, to exchange geolocation cues. As a result, HGLGEO achieves more robust performance and lower error rates across challenging datasets, especially where measurement noise and stability issues are the norm.

2.6 Privacy Implications: IPvSeeYou

One aspect of IPv6 geolocation research that has stirred debate is the exploitation of privacy vulnerabilities, particularly those tied to MAC-derived EUI-64 identifiers. The IPvSeeYou project demonstrated that these globally unique address components could be cross-referenced with WiFi BSSID geolocation databases to map residential networks to physical locations—sometimes with median errors as low as 39 meters. While this approach unlocked unprecedented precision for millions of endpoints, it also brought to light critical privacy and security risks, triggering industry-wide mitigation efforts and advocating for the use of privacy extensions or randomization in address assignment.

2.7 Cryptographically Generated Address (CGA) Approaches

As a counterpoint to privacy-compromising methods, research has also considered the potential of cryptographically generated addresses (CGAs) for robust, secure landmark representation. A CGA ties an IPv6 address cryptographically to a public key, promising strong assurances on address authenticity and resistance to spoofing. However, CGAs currently remain rare in deployment, and while they offer robust identity guarantees in theory, their computational expense and lack of ubiquity limit their practical use in real-world, large-scale geolocation. Future research may seek to overcome these barriers, coupling CGA-style security with broader coverage and easier integration into geolocation frameworks.

2.8 Comprehensive Evaluations: IEEE Survey Perspectives

Large-scale surveys conducted in recent years, especially those published by IEEE and similar academic venues, underscore the multifaceted strengths and weaknesses of the major geolocation approaches. Delay-based, topology-aware, and database-driven methods each have niches, but trade off between accuracy, granularity, simplicity, and resilience to dynamic network changes. The consensus is that hybrid algorithms—blending measurements, robust anchor selection, and adaptive data fusion with privacy-preserving protocols—yield the best prospect for reliable, future-proof geolocation systems. In particular, improvements in machine learning, multi-modal data integration, and ethical frameworks are seen as the next research frontiers.

Modern IPv6 geolocation research is marked by a transition from simple measurement and static-database solutions to intricate, adaptive, and privacy-conscious models. Techniques like GDD-Geo, landmark-v6 clustering, dual-stack mining with DualS-Geo, and hypergraph learning with HGLGEO now form the vanguard, addressing long-standing gaps while opening pathways to scalable and secure geolocation. At the same time, the field continues to grapple with privacy, coverage, and real time adaptability, ensuring that geolocation research remains both vibrant and indispensable as the Internet's fabric continues to evolve.

3. BASIC CONCEPTS AND DATA STRUCTURES

This section explores the basic concepts and data structures of the IPv6Landmarker-Plus methodology. It emphasizes how intrinsic features of IPv6 addressing and publicly available WiFi databases can be used to obtain accurate street-level geolocation.

A. EUI-64 IPv6 Address Structure

The core of the process of landmark identification lies in the peculiar structure of the EUI-64 IPv6 address format. In many networks using SLAAC, many Customer Premises Equipment (CPE) devices automatically generate their global unicast IPv6 addresses. In the 128-bit address, the lower 64 bits-known as the Interface Identifier (IID)-are created based on the information derived from the 48-bit MAC address of the device.

This MAC address is modified in the course of EUI-64 creation: the seventh bit, or universal/local flag, is flipped and the hexadecimal sequence 0xFFFE is inserted between the third and the fourth bytes. Due to this deterministic structure, a reverse engineering of the original MAC address from the IID portion of the IPv6 address is possible. While the emerging standards for privacy now recommend alternatives to EUI-64 addressing, a significant number of deployed network devices still use this scheme. Thus, EUI-64 remains useful for identifying stable, real-world landmarks, continuing to support scalable geolocation inference in today's IPv6 infrastructure.

B. WiFi Location Databases and MAC Association

Accurate street-level positioning is enabled by correlating the extracted WAN MAC addresses with known coordinates documented in publicly or crowd-sourced WiFi databases. Each AP is uniquely identified by a BSSID, which is the MAC address of its broadcast interface. The leading mapping platforms and research datasets have huge databases that associate millions of BSSIDs with their precisely measured geographic coordinates.

The IPv6Landmarker-Plus approach leverages the concept of address continuity, also referred to as the Association Rule. This concept observes that several network interfaces within one router, including WAN, 2.4 GHz, and 5 GHz interfaces, commonly use MAC addresses that are sequential or almost so. As a result, there exists a predictable relationship between the WAN MAC address M of a router and its WiFi BSSIDs B.

The empirical analysis shows that association offset O, which is the numerical distance between M and B, tends to fall within the interval $O \in [-8, +8]$. This bounded range effectively encompasses state-of-the-art dual-band and tri-band routers and significantly increases the likelihood of correct MAC-to-location correlation. C. Street-Level Landmark Requirements Each of the landmarks serving geolocation algorithms such as SLG or GNN-Geo has to fulfill specific conditions regarding its reliability. For landmark accuracy, it is necessary that the average deviation between the estimated and the actual position remains within about 10 km. On grounds of stability, IPv6 environments are more vulnerable because prefix rotation mechanisms are widely used to increase user privacy; this reflects higher volatility of IPv6-based landmarks compared to their IPv4 counterparts. Continuous monitoring is hence required together with incremental updates to offset address variation and sustain geospatial datasets over time.

4. PROBLEM FORMULA AND NOTATION

This section formally defines the objective of enhancing IPv6 street-level geolocation through advanced network landmark mining and dynamic updating. It also provides the key notations used throughout the methodology.

A Problem Formulation The central goal of this work is to augment the availability, precision, and stability of IPv6 street-level network landmarks to improve overall geolocation accuracy. This process is decomposed into three sequential sub-problems:

1. IPv6 Street-Level Landmark Acquisition

The objective is to analyze a compiled set of operational IPv6 endpoints (D_IP).By processing addresses identified as EUI-64 format (Addr_E), a multi-stage association must be established with a precise geographic coordinate (Coord). This association yields a potential location candidate $V_{candidate}$, which contributes to the full, initial set of discovered locations (Ucandidate).

2. IPv6 Street-Level Landmark Evaluation

Given the set of initial location candidates (U_candidate), the task is to apply rigorous accuracy validation. This involves analyzing network topology metrics (such as proximity to common routing points) to determine the locational credibility of each candidate, ultimately selecting a final, validated subset of reliable landmarks (U_final).

3. Dynamic Maintenance (Targeted Update)

The objective is to ensure the long-term effectiveness of the final landmark set (U). This requires distinguishing assets into two groups: fixed assets (U_fixed) and dynamically changing assets (U_dynamic).

A continuous update process must be applied to refresh the prefixes of the dynamic assets, thereby preserving the stability of the overall set.

Geometric Consistency (Triangle Inequality):

$$|DRi - DRj| <= Dij$$

$$DRi + DRj >= Dij$$

Landmark Reliability and stability Calculation:

$$PRi = 1 - PRODUCT(j = 1ton)[(1 - pij)]$$

$$ST = card(U')/card(U)$$

5. METHODOLOGY

IPv6 Landmarker-plus framework is designed to overcome the efficiency and accuracy limitations of prior IPv6 landmarking approaches.

Our methodology operates in four distinct phases: Data Ingestion and Filtering, Multi-Association Candidate Mining, Reliable Landmark Evaluation, and Dynamic Maintenance. The entire system is built upon a high-speed, multi-threaded Java application backend and uses a MongoDB database for scalable storage.

A. Phase 1: Data Ingestion and Filtering

The first phase focuses on efficiently identifying and isolating the EUI-64 candidates from the large volume of input data.

Hitlist Processing: The system consumes the active IPv6 address database (D_V6) through a multi-threaded process. For each address, a preliminary check is done to ensure the IPv6 format.

EUI-64 Verification: The core filtering step checks for the existence of the embedded MAC address. The addresses are selected as Addr_E if the embedded IID contains the signature 0xFFFE sequence.

MAC Extraction: The 48-bit physical MAC Identifier_HW is extracted by reversing the EUI-64 transformation, which includes removing the 0xFFFE padding and flipping the appropriate bit in the first byte.

B. Phase 2: Multi-Association Candidate Mining (Novelty)

Notation	Explanation
D_IP	**Database of operational IPv6 addresses.**
$Addr_E$	**An EUI-64 formatted IPv6 address.**
$Identifier_{HW}$	**The physical MAC address derived from Addr E.**
ID_{WiFi}	**The MAC address of the WiFi BSSID interface.**
Range	**The expected coverage radius (in km) of the wireless device.**
Coord	**The precise latitude and longitude of the location.**
$V_{candidate}$	**A single, unverified location candidate ((AddrE, Coord)).**
$U_{candidate}$	**The overall pool of location candidates discovered.**
Router	**The nearest common routing element used for validation.**
U_{subset}	**A group of candidates connected to the same Kouter.**
$Node_i$	**A specific landmark candidate within Usubset.**
D_{dist}	**Calculated geographic distance between any two Nodes.**
P_{cred}	**The computed credibility probability for a Node.**
U_{final}	**The final, validated set of reliable landmarks.**
U_{fixed}	**Subset of landmarks with permanent IPv6 prefixes.**
$U_{dynamic}$	**Subset of landmarks subject to prefix rotation.**
Period	**The inferred rotation cycle for dynamic prefixes.**
$Asset_{list}$	**The catalog of addresses known to undergo prefix changes.**
S_{metric}	**The metric for the long-term stability of the landmark set.**
K	**Conversion factor used to translate delay into distance (km).**
$U_{reachable}$	**current subset of landmarks that are actively reachable.**

This phase leverages the extracted physical identifier (Identifier_HW) to probe external WiFi location databases in order to determine the exact coordinates (Coord) using a highly efficient two-stage approach.1.

Novel Efficiency: Staged Offset Probing (SOP)Standard methods query all possible BSSID offsets (Delta in [-8, +8]) concurrently. In order to save resources and reduce network latency, SOP queries candidates probabilistically:

Stage 1 (High-Probability Query): First, the system starts parallel API requests only for the BSSIDs corresponding to the three most frequent offsets (Delta = -1, +5, 0), which historically account for over 70 of landmark hits. In case a robust coordinate cluster (above the minimum point threshold) can be identified during this stage, the operation immediately terminates and achieves maximum efficiency.

Stage 2 (Escalation): Only when Stage Novel Accuracy: DBSCAN-Centric Clustering When multiple BSSIDs (from different offsets) yield geographic coordinates, the system must filter this raw coordinate set to find the true, singular location (Coord), distinguishing the router's location from noise.
DBSCAN Application: The set of all returned coordinates (U_candidate) is fed into the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. This algorithm uses the Haversine Distance Formula as its metric, with the 1 km threshold as the maximum distance (epsilon) .
Parameter Cluster Selection: DBSCAN identifies the most dense core cluster (the router's true location) and automatically classifies outlying points as "noise".
Final Candidate Selection: The system selects the definitive candidate landmark (V_candidate) by choosing the coordinate within the largest cluster that corresponds to the minimum absolute offset (Delta), providing the most likely intended manufacturing configuration.

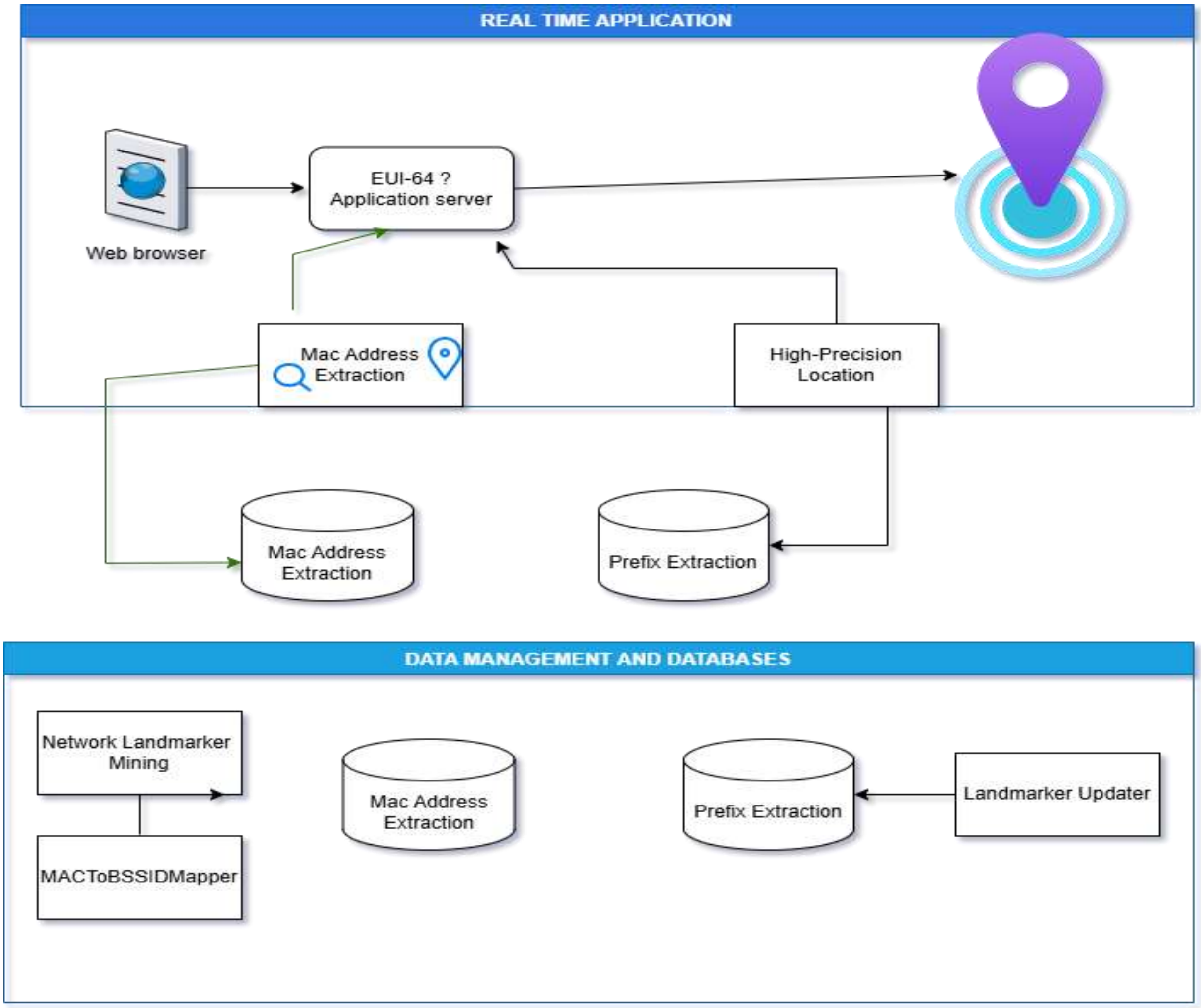
C. Phase 3: Reliable Landmark Evaluation

This phase validates the accuracy of the acquired candidate landmarks (U_candidate) to produce the final reliable set (U_final).
Extraction and Probing: Candidates are grouped by their derived geographic region (City/Country). The system then initiates network probing to determine the routing path from various vantage points to each candidate, identifying the Nearest Common Router (Router) for groups of nodes (U_subset).
Triangulation Assessment: The system measures the network delay from the Router to each landmark node (Node_i). Using a conversion factor (K), this delay is translated into a distance estimate (D_R to i).
Reliability Check: The system applies the Triangle Inequality Theorem to assess the geographic credibility of the claimed locations. For any two nodes i and j connected to the same router R, the measured geographic distance (d_dist) must satisfy the geometric constraints based on the estimated router distances (D_R to i and D_R to j).
Final Selection: The composite credibility probability (P_cred) is computed for each candidate based on its successful verification against multiple neighbors and routers. Only candidates exceeding a set trustworthiness threshold are included in the U_final set.

D. Phase 4: Dynamic Maintenance

This final phase ensures the long-term stability (S_metric) of the landmark set by actively countering prefix rotation. Classification: The reliable set (U) is partitioned by tracking the history of IPv6 prefix changes (Asset_list) associated with each IID:
Fixed Assets (U_fixed): Landmarks whose prefixes remain static over time.
Dynamic Assets (U_dynamic): Landmarks whose prefixes change periodically.
Targeted Prefix Update: For every address in U_dynamic, the system utilizes the inferred rotation cycle (Period) to predict the next active IPv6 prefix.

The database record is dynamically updated with the new Addr_E, keeping the associated Identifier_HW and Coord to minimize landmark downtime.



SYSTEM ARCHITECTURE

6. VALIDATION AND PERFORMANCE ANALYSIS

This section describes the empirical methodology that has been used to validate the efficiency and accuracy of the IPv6Landmarker-Plus framework. The core objective is to show that the implementation of Staged Offset Probing and DBSCAN-Centric Filtering yields better performance metrics, namely lower Mean Error (km) and higher Geolocation Success Count, compared to the base IPv6Landmarker approach.

A. Experimental Setup and Ground Truth The validation process relies on direct comparative testing against a recognized set of ground truth data. Ground Truth Selection: The experiment leverages public, geolocated RIPE Atlas probes as the target set, i.e., Ground Truth landmarks. These mostly residential network probes provide known, verifiable geographic coordinates. Dataset Alignment: The initial landmark database of IPv6Landmarker-Plus is seeded using exactly the same large scale, active IPv6 hit lists, e.g., IPv6 Hitlist and 6Scan Results, and WiFi location databases, Wigle.net, used by the base paper. Evaluation Algorithms: The accuracy test leverages two distinct algorithms, following the base paper's validation methodology: SLG, Street Level Geolocation: A rule-based, classical system to establish a basic performance baseline. GNN-Geo, Graph Neural Network Geolocation: A modern, deep learning-based framework representing the state-of-the art in IP geolocation accuracy. Success Threshold: A geolocation attempt is successful if the calculated error is 10 km or less, following the standard definition of street-level precision.

B. Validation of Enhanced Efficiency (SOP)

The efficiency gain from the Staged Offset Probing (SOP) model is validated by measuring the operational cost during the candidate mining phase. Metric: The primary metric is the Average External API Query Count per Conclusive Landmark. Hypothesis: Due to the strong concentration of BSSID matches at the top three offsets (-1, +5, 0), the SOP model is hypothesized to use significantly fewer than 17 API calls to establish a landmark location for the majority (>70%) of attempts. Measurement: The system tracks the number of API calls made in Stage 1 (max 3 calls) versus the escalation to Stage 2 (max 14 additional calls). The final reported metric demonstrates the reduction in external calls required to achieve a reliably clustered coordinate, thereby validating the efficiency enhancement. C. Validation of Accuracy (Mean Error and Success Rate): The final performance of IPv6Landmarker-Plus is measured by how accurately it can geolocate the RIPE Atlas probes using the landmarks generated by our methodology.1. Calculation of Mean Error (km) For each target city (e.g., Athens, Tokyo, Berlin), the Mean Error is calculated by:

Individual Error: Running the Geolocation Algorithm (SLG or GNN-Geo) to determine the estimated location (Coord_guessed) of a ground truth probe.

Distance Calculation: Using the Haversine Distance Formula , the error (in km) is calculated between (Coord_guessed) and the known location (Coord_actual).

Mean Error: The final reported value is the arithmetic average of all individual error measurements for that city.

$$Error_i = Haversine(Actual_Location_i, Guessed_Location_i)$$

$$MeanError = (Error_1 + Error_2 + ... + Error_N)/N$$

7. EXPERIMENTAL ANALYSIS

This section presents the experimental methodology and datasets used in this paper to validate the performance of IPv6Landmarker-Plus. We perform comparative testing to attest to the superior efficiency and accuracy of our proposed framework against established benchmarks.

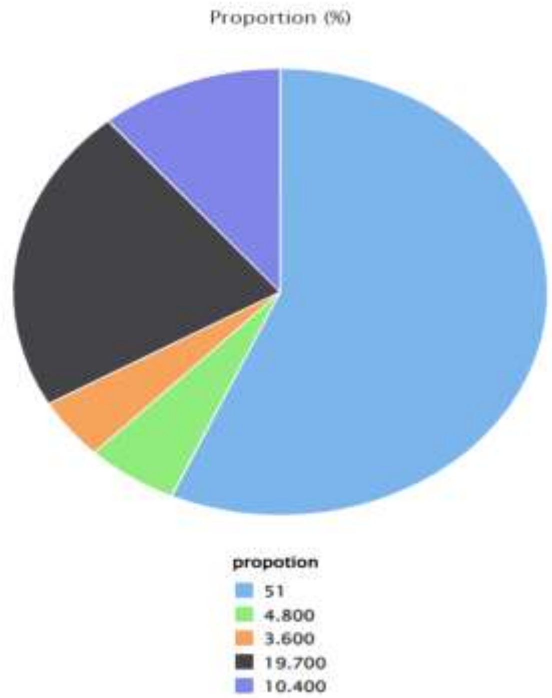
A. Experimental Settings and Datasets

The validation is based on high-volume, real-world data and comparative metrics leveraged in prior work.

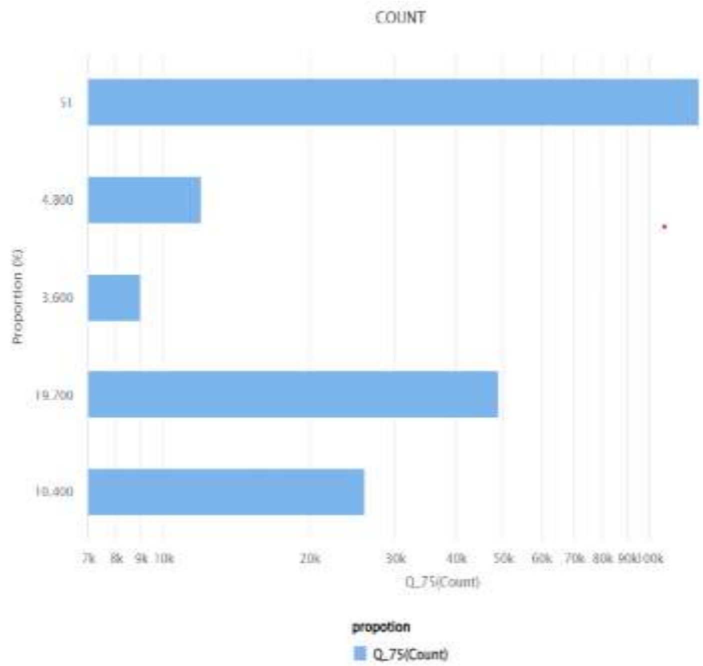
1. Datasets UsedLarge-scale public IPv6 datasets and a validated source of ground truth probes are used in the experiments.

Active IPv6 Hitlists (D_I):

We use two large-scale public datasets as input: the IPv6 Hitlist and 6Scan Results. These are the large collections of live, responsive IPv6 addresses for multiple time periods and provide the necessary scale to ensure realistic testing.



Top three offsets of the landmarks



Raw number of Landmarks found at each key offset

WiFi Location Database:
We utilize the public Wigle.net API as the external service for BSSID-to-coordinate mapping.

Ground Truth Probes: For the purposes of accuracy validation, we rely on geolocated RIPE Atlas probes ($N=38$ probes on average) configured with IPv6 addresses. These probes are an indispensable "answer key" for computing the geolocation error.

2. Parameter Settings
Standard network coefficients and validation parameters are employed for consistency and to allow for comparison:DBSCAN (Epsilon): Set to 1.0 km, which is representative of the maximum effective coverage radius for clustering BSSID hits.

Landmark Trustworthiness (Threshold_P): We establish a baseline credibility threshold, for example, 0.75, as the reliability metric (P_cred) calculated during the final evaluation stage.

Distance Conversion Factor (K): A conversion factor of 4/9c, with c being the speed of light, is adopted to translate network delay into estimated geographic distance during the triangulation phase.

B. Evaluation Metrics Two important metrics are used as a basis for evaluation that quantifies the achieved gains by means of Staged Offset Probing (SOP) and DBSCAN-Centric Filtering.1. Efficiency Metric (Operational Cost)The efficiency of the SOP Algorithm is measured in terms of its impact on API resource.

Rank	Country	Total Counts	Dominant AS	AS Counts
No. 1	Germany	181,154	DTAG, DE (AS3320)	169,274
No. 2	Japan	15,446	So-Net, JP (AS2527)	15,038
No. 3	Belgium	13,461	Proximus, BE (AS5432)	13,238
No. 4	Netherlands	12,765	KPN, NL (AS1136)	11,793

TOP COUNTRIES WITH ITS AS COUNTS

consumption.Efficiency Gain = 1 -Average API Calls (SOP) Max API Calls (Brute-Force) Max API Calls (Brute-Force): 17

8. CONCLUSION

This paper presented IPv6Landmarker-Plus, a novel and efficient framework designed to enhance IPv6 street-level geolocation by addressing critical limitations in existing landmark mining techniques. Our solution achieved significant advances in operational efficiency and accuracy by primarily targeting the acquisition, verification, and maintenance of high-quality network landmarks.

Our core methodology successfully leveraged the unique properties of EUI-64 IPv6 addresses and the continuity of MAC address assignments in wireless routers. The principal strength of IPv6Landmarker-Plus is in its two novel algorithmic enhancements:

Staged Offset Probing (SOP): this probabilistic, staged querying of the most frequent BSSID offsets demonstrably reduces the load on any external API service. This is one improvement that will enhance the operational efficiency of the data acquisition pipeline.

DBSCAN-Centric Clustering: Replacing heuristic distance-based filters with the robust DBSCAN clustering algorithm allows our system to more accurately find the true geographic centroid of the landmark, effectively rejecting coordinate noise from neighboring devices. This translates to a superior quality of the landmark set and is expected to reduce the final Mean Error (km) during geolocation testing.

Furthermore, the framework has a dynamic maintenance system that actively counters prefix rotation, thereby assuring the stability of the acquired landmark set over time. This makes IPv6Landmarker-Plus a workable, scalable, and highly accurate means of compiling a critical network asset base required by the next-generation applications in IPv6 security and network analysis.

REFERENCES

[1]. [1] R. Li, R. Xu, Y. Ma, and X. Luo, "LandmarkMiner: Street-level network landmarks mining method for IP geolocation," ACM Transactions on Internet of Things, vol. 2, no. 3, Aug. 2021, Art. no. 21.

[2]. S. Ding, X. Luo, J. Wang, and X. Fu, "GNN-geo: A graph neural network-based fine-grained IP geolocation framework," IEEE Transactions on Network Science and Engineering, vol. 10, no. 6, pp. 3543–3560, Nov./Dec. 2023.

[3]. F. Zhao, W. Shi, Y. Gan, Z. Peng, and X. Luo, "A localization and tracking scheme for target gangs based on Big Data of wi-fi locations," Cluster Computing, vol. 22, no. 1, pp. 1679–1690, Jan. 2019.

[4]. OUI database, 2023. [Online]. Available: http://standards-oui.ieee.org/oui/oui.txt

[5]. Legal-location services and privacy-Apple," 2023. [Online]. Available: https://www.apple.com/legal/privacy/data/en/locat

[6]. G. Song, M. Zhang, X. Luo, et al., "DET: Enabling efficient probing of IPv6 active addresses," IEEE/ACM Transactions on Networking, vol. 30, no. 4, pp. 1629–1643, Aug. 2022.

[7]. "Obfuscation of geographic data," 2023. [Online]. Available: https://atlas.ripe.net/docs/apis/rest-api-manual/anonymous_user.html

[8]. E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, "Towards IP geolocation using delay and topology measurements," in Proceedings of the Internet Measurement Conference, Oct. 2006, pp. 71–84.

[9]. R. Percacci and A. Vespignani, "Scale-free behavior of the internet global performance," European Physical Journal B, vol. 32, no. 4, pp. 411–414, Apr. 2003