

Galeo: Continuous Bayesian Topology-Aware IP Geolocation

Leo Galambos

Abstract—This paper archives and formalizes *Galeo*, an unpublished Internet Protocol (IP) geolocation method developed by the author approximately fifteen years before this publication and reconstructed here for archival purposes. Galeo treats a corpus of distributed traceroutes as a coupled spatial inference problem rather than converting only end-to-end delay to a target radius. Measurement probes have known coordinates; intermediate routers are persistent latent spatial variables shared by all traceroutes in which they occur; selected routers may be geographic anchors; and the target IP address is localized by marginalizing uncertainty over the whole measurement graph. Delay observations are modeled by calibrated generative delay-distance likelihoods; when expressed in range space they induce annular likelihoods or radial potentials in two dimensions. A mesh of synchronized probes can additionally provide directed one-way delay (OWD) calibration. A soft geographic prior favors settlements and plausible network infrastructure while retaining non-zero probability for rural or exceptional installations. Two regimes are distinguished: a local or administratively controlled network in which path symmetry has been verified, and the general Internet where reverse paths and routing states must be treated as latent variables. A retrospective review shows substantial independent prior art for individual components: TBG simultaneously localizes routers and targets; Octant performs transitive intermediate-router localization and uses population/ocean constraints; Laki *et al.* combine topology, a detailed path-latency model, and GPS-synchronized OWD; Spotter uses a normal delay-conditioned distance model to form probabilistic annuli for the final target; and later factor-graph and common-router methods introduce joint probabilistic or secondary-landmark inference. The literature examined through September 2026 did not reveal one publication combining all of Galeo’s continuous latent-router posterior, shared-router coupling, annular edge factors, OWD calibration, route-state marginalization, router anchors, and geographic/infrastructure priors. This observation is a retrospective positioning of the archived method, not a claim of chronological priority or patent novelty. Synthetic examples with 24 and 36 probes illustrate the formulation, including the sharpening obtained when intermediate routers are promoted to probabilistic secondary landmarks, and its uncertainty behavior.

Index Terms—IP geolocation, Galeo, traceroute, Bayesian inference, factor graph, router geolocation, one-way delay, topology, spatial prior, route asymmetry.

I. INTRODUCTION AND RETROSPECTIVE PRIOR ART

A. Archival Scope

Inferring the physical position of an Internet host from active measurements is intrinsically underdetermined. Internet routing is policy-driven rather than geographic; paths can be asymmetric; queueing and processing delays contaminate propagation delay; load balancing can expose multiple paths; and physical cable length exceeds straight-line distance. Consequently, a single conversion such as “delay multiplied by propagation speed” cannot, in general, identify a geographic position.

This paper documents *Galeo*, a method developed by the author in unpublished form approximately fifteen years before this archival publication. The original idea was to exploit a serious distributed measurement apparatus—at least lower tens of probes—that repeatedly traceroutes both between probes and toward an unknown address X . Known probe coordinates anchor the topology. Each observed router is a persistent spatial variable; when routes intersect, the same router variable couples the measurements. A measured delay does not determine one distance; through the calibrated measurement model it induces a distribution over range and, in two dimensions, an annular likelihood or radial potential. The distribution is propagated through successive hops and combined with messages arriving through other traceroutes. Known router coordinates sharpen the graph. Calibration paths between probes constrain the delay-distance relation. The final output is a posterior density for X , rather than only a point estimate. Crucially, a discovered router need not remain merely a passive hop: once recognized on traceroute paths and directly probed from additional probes, it becomes a *probabilistic secondary landmark*. Its sharpened posterior can then sharpen downstream routers and, ultimately, the posterior of X .

The method is reconstructed here using modern probabilistic graphical-model notation. This is intentionally an *archival description*, not an assertion that its components were unknown when Galeo was conceived. A retrospective literature review was therefore performed against the closest measurement-based, topology-aware, probabilistic, and graph-based IP-geolocation methods, supplemented by a recent broad survey [27]. The purpose is to state precisely what Galeo shares with prior work and what combination of mechanisms characterizes the archived method.

B. Closest Precedents

Early active geolocation established the basic latency/distance paradigm [8]. Constraint-Based Geolocation (CBG) self-calibrates delay-to-distance bounds and multilaterates a target in continuous geographic space [9]. GeoBuD subsequently used traceroute-derived buffering-delay estimates at intermediate hops to improve those constraints [10].

Topology-Based Geolocation (TBG) is a particularly close precedent. Katz-Bassett *et al.* convert route topology and delay into constraints on unknown target and intermediate-router locations and solve for routers and targets simultaneously [11]. The coupling of shared routers and the requirement for a globally consistent map are therefore not unique to Galeo.

Measurement graph: probes, shared routers, anchors, and target

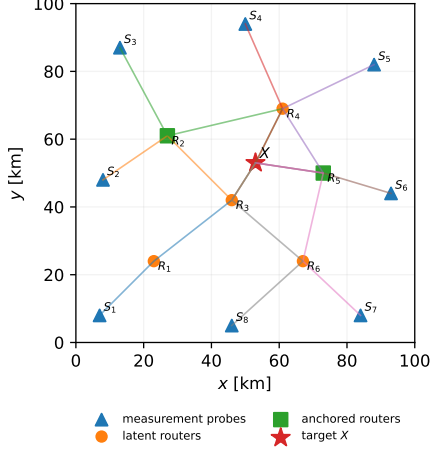


Fig. 1. Galeo measurement graph. Triangles are probes with known coordinates, circles are latent routers, squares are routers with external geographic anchors, and X is the target. A router observed on several paths is represented once and couples all corresponding measurements.

The principal distinction is representational: TBG solves a constraint/global-optimization problem for a best map, whereas Galeo retains continuous probability densities for latent router coordinates and propagates/marginalizes their uncertainty.

Octant is another close predecessor. It represents positive and negative constraints as weighted geographic regions, performs transitive localization by using intermediate nodes as secondary landmarks, and experimentally applies population-density and ocean constraints [12]. Thus neither “localize through intermediate routers” nor “suppress implausible geography” should be presented as novel Galeo components. Galeo differs in representing hop uncertainty as probabilistic spatial factors inside one coupled continuous posterior rather than as geometric constraint regions.

Youn *et al.* estimate statistical delay-distance relationships with kernel density estimation and maximize target likelihood [13]. Laki *et al.* are even closer to Galeo’s calibration concept: their path-latency model uses discovered topology, per-link delay reasoning, global optimization, and precise OWD measurements from the GPS-synchronized 18-node ETOMIC measurement infrastructure [14], [15]. Spotter subsequently fits the conditional distance distribution for a measured delay with a normal model, maps it to a ring-like spatial density around each landmark, and multiplies multiple landmark densities to obtain a probabilistic target map [16]. Spotter explicitly notes that population or geographic information can enter the prior, although its reported experiments use a uniform prior. Spotter, however, does not make every intermediate router a latent spatial variable in the target inference graph.

Later work narrows the gap further. Chandekar and Paris automatically construct factor graphs and use sum-product inference to geolocate many network hosts simultaneously, but locations are discrete geographic regions rather than continuous coordinates [17]; Laskey *et al.* encode the same family of

reasoning in a probabilistic ontology/Bayesian-network representation [18]. Landmark-calibration, nearest-common-router, and Common-Routers-based Geolocation (CRG) methods locate intermediate/common routers and reuse them as secondary landmarks [19]–[21]. CRG even learns a truncated-normal $p(d | RTT, N)$ model, but explicitly abstracts detailed topology between common routers and end nodes by the hop-count variable N , rather than maintaining all routers in one detailed continuous latent graph [21]. For Internet interconnection mapping, mi² uses localized traceroutes and belief propagation on a Markov random field to assign interconnections to known colocation facilities [22]; this is an important graph-propagation precedent, but its location state is facility membership rather than a continuous posterior over arbitrary router coordinates. Recent GraphGeo, GNN-Geo, and TRFGeo methods exploit topology by graph neural networks and learned message passing [23], [24], [26]; these are data-driven graph-regression systems rather than the generative physical/probabilistic model considered here.

C. What Is Specific to the Galeo Synthesis

Within the primary literature reviewed and the 2025 survey [27], no single publication was identified that simultaneously combines all of the following: (i) continuous spatial PDFs for every shared intermediate router; (ii) one persistent router variable across intersecting traceroutes; (iii) learned annular hop likelihoods; (iv) synchronized probe-to-probe OWD calibration; (v) hard or uncertain geographic router anchors; (vi) explicit latent forward/reverse route states when symmetry cannot be guaranteed; and (vii) soft settlement/infrastructure priors applied throughout the latent graph. This is a qualified literature observation, not a proof of priority: the search cannot exclude unpublished work, patents outside the reviewed corpus, or literature using different terminology.

The mathematical target of Galeo is the normalized posterior

$$p_X(x, y | \mathcal{D}, \mathcal{C}), \quad (1)$$

obtained by marginalizing router coordinates, calibration parameters, and route states. The remainder of the paper defines this model.

II. MEASUREMENT SEMANTICS AND ASSUMPTIONS

A. Traceroute Does Not Directly Measure One-Way Hop Delay

Standard traceroute discovers a sequence of responding interfaces on the forward path, but the time reported for each response is a round-trip time (RTT) from the initiator to that responding hop and back [1]. If probe S_i elicits an ICMP Time Exceeded response associated with router R_k , the observed time is more accurately represented as

$$T_{ik}^{\text{RTT}} = F_{i \rightarrow k} + R_{k \rightarrow i} + c_{ik} + \eta_{ik}, \quad (2)$$

where $F_{i \rightarrow k}$ is forward-path delay, $R_{k \rightarrow i}$ is the ICMP return-path delay, c_{ik} is responder/control-plane generation and scheduling delay, and η_{ik} collects residual timestamping and queueing variation. Router requirements explicitly permit rate limiting of ICMP error messages, and the ICMP source address

TABLE I

RETROSPECTIVE POSITIONING OF GALEO AGAINST THE CLOSEST VERIFIED PRIOR WORK. “JOINT” MEANS THAT UNCERTAINTY OR COORDINATES OF MULTIPLE NETWORK NODES ARE COUPLED IN ONE INFERENCE PROBLEM; “CONTINUOUS POSTERIOR” MEANS AN EXPLICIT PROBABILITY DENSITY OVER CONTINUOUS SPATIAL COORDINATES, NOT ONLY A BEST-FIT POINT OR WEIGHTED FEASIBLE REGION.

Method	Delay/distance representation	Intermediate routers	Inference/output	Geographic/context treatment
CBG (2006) [9]	Self-calibrated delay-derived distance bounds	Not persistent inference variables	Continuous target constraint region / multilateration	None central
GeoBuD (2006) [10]	Per-hop buffering-delay correction from traceroute	Used to estimate buffering effects	Improved CBG target constraints	None central
TBG (2006) [11]	Topology + delay constraints	All encountered routers ; target and routers solved simultaneously	Global constraint optimization; best-fit map	Validated passive reference-node hints
Octant (2007) [12]	Weighted positive/negative geometric constraints	Transitive secondary landmarks from intermediate nodes	Error-tolerant weighted feasible regions	Population-density and ocean constraints
Youn <i>et al.</i> (2009) [13]	KDE statistical delay-distance density	No detailed latent router graph	Maximum-likelihood target point	Not central
Laki <i>et al.</i> (2009/10) [14], [15]	Detailed path-latency model + precise OWD	Routers along measured paths	Global optimization of router geometry	GPS-synchronized probe anchors
Spotter (2011) [16]	Normal $p(d \mid \text{delay})$; ring-like spatial PDFs	Not latent variables in target inference	Product of landmark PDFs; probabilistic target map	Prior can encode population/geography; experiments use uniform prior
Factor graph (2015) [17]	Probabilistic delay evidence	Arbitrary hosts coupled by links	Joint sum-product ; discrete region PMFs	Multi-factor evidence; discrete geography
Landmark calibration / NCR / CRG (2016–19) [19]–[21]	Relative delay; CRG truncated-normal $p(d \mid RTT, N)$	Common routers localized as secondary landmarks	Staged/optimization/ML localization; detailed router-to-router topology abstracted	Region-aware parameters
mi ² (2019) [22]	Latency/topology and co-presence evidence	Border routers/interconnections on localized traceroutes	Belief propagation on an MRF; discrete colocation-facility state	Facility membership, reverse-DNS/known-location anchors
GraphGeo/GNN-Geo/TRFGeo (2022–26) [23], [24], [26]	Learned network/attribute embeddings	Nodes linked in learned/observed graph	GNN regression/message passing	Rich learned features; not a physical posterior
Galeo (archived here)	Learned Gaussian/robust hop factors; OWD-calibrated path factors	All shared routers are persistent continuous latent variables	Joint continuous posterior , message passing and marginalization over routers/routes	Soft settlement/infrastructure priors ; anchored routers; explicit route-state uncertainty

need not identify the ingress interface that received the probe [2]. Thus a traceroute RTT is not a transparent measurement of data-plane propagation alone.

For adjacent forward hops R_k and R_{k+1} ,

$$\begin{aligned} T_{i,k+1}^{\text{RTT}} - T_{ik}^{\text{RTT}} \\ = d_{k,k+1}^+ + (R_{k+1 \rightarrow i} - R_{k \rightarrow i}) + (c_{i,k+1} - c_{ik}) + \Delta\eta, \end{aligned} \quad (3)$$

not simply the delay of the forward edge. Negative incremental RTTs are therefore possible and should be treated as evidence of return-path, responder-delay, or transient effects rather than as negative link distance. This distinction is fundamental in the wide-area model.

B. One-Way Probe-to-Probe Measurements

If both endpoints are controlled probes and their clocks are synchronized sufficiently well, the probes can measure a true one-way delay (OWD). The IPPM one-way delay metric explicitly separates the source-to-destination direction from the reverse direction [3], and OWAMP provides a protocol for active one-way measurement [4]. For a packet sent at source time t_s and received at destination time t_r ,

$$Z_{ij} = t_r - t_s = \tau(\Pi_{i \rightarrow j}) + \epsilon_{ij} + \epsilon_{\text{clk}}, \quad (4)$$

where $\Pi_{i \rightarrow j}$ denotes the directed path and ϵ_{clk} is residual clock synchronization error.

One-way probe measurements are high-value calibration observations because the return path does not enter (4). They do not, by themselves, identify the delay of each constituent router-to-router segment. Per-segment inference

becomes possible only to the extent that overlapping paths, anchors, repeated measurements, and pooled delay parameters make the decomposition observable.

C. Path Stability and Load Balancing

Traceroute probes should preserve flow-identifying header fields so that classic per-flow equal-cost multi-path (ECMP) forwarding does not create an artificial path assembled from different branches. Paris traceroute was designed to reduce this ambiguity [5]. It is a mitigation, not a guarantee: per-packet load balancing, route changes, stateful middleboxes, and ICMP return paths can still vary. Galeo therefore associates each measurement with an observed path family or a distribution over supported path hypotheses; path changes are not silently averaged into one deterministic sequence.

D. Interface Identity, Aliasing, and Hidden Hops

Traceroute exposes responding IP interfaces, not physical routers. Let $\mathcal{I}$ be the set of observed interface addresses and let

$$a : \mathcal{I} \rightarrow \mathcal{R} \quad (5)$$

map interfaces to router identities. When alias resolution is reliable, a may be fixed during graph construction; otherwise it should be treated as uncertain and propagated as an additional latent discrete variable. Internet-scale alias-resolution systems such as MIDAR illustrate why this step is necessary for router-level topology [6]. Without it, two interfaces of one router may be mistaken for two spatially distinct routers, while an ICMP source address may also differ from the interface on which the probe arrived [2].

Likewise, an observed IP-layer adjacency need not be a single physical link. MPLS tunnels, layer-2 transport, non-responding routers, and filtering can hide forwarding elements; work such as TNT demonstrates that traceroute-visible topology can omit MPLS structure [7]. Galeo consequently interprets an edge as a measured logical segment. A segment may represent one physical link or an unresolved sequence of links, with the latter absorbed into its latent path-length and delay parameters.

III. SPATIAL AND GRAPH MODEL

After interface-to-router reconciliation, let the directed measurement graph be

$$G = (V, \mathcal{E}), \quad V = \mathcal{S} \cup \mathcal{R} \cup \{X\}, \quad (6)$$

where $\mathcal{S} = \{S_1, \dots, S_n\}$ is the probe set, $\mathcal{R}$ the router set, and X the unknown target. Each node v has a two-dimensional coordinate

$$\mathbf{q}_v = (x_v, y_v) \in \mathbb{R}^2. \quad (7)$$

The planar model is appropriate for regional examples. Continental or global implementations should use a geodesic metric on an ellipsoid or an Earth-centered coordinate system with an explicit surface-area measure.

For a logical segment $e = (u, v)$ define the endpoint separation

$$d_e(Q) = \|\mathbf{q}_u - \mathbf{q}_v\|_2. \quad (8)$$

The routed physical length need not equal d_e . Introduce a path-stretch variable $\lambda_e \geq 1$,

$$\ell_e = \lambda_e d_e. \quad (9)$$

A baseline delay model for repetition r is

$$\delta_e^{(r)} = b_e + \beta_e d_e(Q) + q_e^{(r)} + \varepsilon_e^{(r)}, \quad (10)$$

where $\beta_e = \lambda_e/v_e$ combines path stretch and effective propagation speed, b_e captures fixed transmission/processing components not represented elsewhere, $q_e^{(r)} \geq 0$ is variable queuing, and $\varepsilon_e^{(r)}$ is residual measurement error. The decomposition is a model of a logical segment, not a claim that traceroute exposes every physical link.

Latency alone cannot separately identify λ_e and v_e , and a completely free β_e for every edge would be nearly confounded with the unknown geometry. Galeo therefore requires pooled or hierarchical calibration. For an edge class $c(e)$ (for example a region, medium, provider/path family, or a deliberately coarse unknown class), one convenient hierarchy is

$$\log \beta_e \sim \mathcal{N}(\mu_{\beta, c(e)}, \tau_{\beta, c(e)}^2), \quad (11)$$

$$q_e^{(r)} \sim F_{q, c(e)}^+, \quad (12)$$

where F^+ is a non-negative queuing distribution such as an exponential/lognormal family or a robust positive mixture. Small $\tau_{\beta, c}$ enforces shrinkage toward a calibrated class relation; larger values are justified only where the probe mesh provides enough independent information. This hierarchy makes explicit that geometry is inferred from shared structure and calibration, not by assigning every segment an arbitrary private speed.

IV. PRIORS, ROUTER ANCHORS, AND GEOGRAPHIC PLAUSIBILITY

Probe coordinates are fixed:

$$p(\mathbf{q}_{S_i}) = \delta(\mathbf{q}_{S_i} - \mathbf{q}_{S_i}^*). \quad (13)$$

A router whose position is known precisely can be treated identically. If an external source provides an uncertain router location $\mathbf{q}_r^*$ with covariance Σ_r ,

$$p(\mathbf{q}_r) = \mathcal{N}(\mathbf{q}_r; \mathbf{q}_r^*, \Sigma_r). \quad (14)$$

These anchor factors can represent operator knowledge, reliable hostname-derived locations, known points of presence, exchange facilities, or other independently validated evidence. They must not be double-counted if the same source also appears in calibration data.

A. Settlement and Infrastructure Prior

A purely geometric annulus assigns the same probability to a lake, an agricultural field, a city center, and a carrier hotel at equal radius. That is physically implausible, but replacing the annulus by a hard “city-only” mask is also wrong: routers can occur in rural cabinets, mobile sites, industrial areas, cable stations, roadside facilities, or other infrastructure outside settlement centers. Galeo therefore uses a *soft* spatial prior.

Let settlement catalogue entry j have centroid $\mathbf{c}_j$, population or importance weight N_j , and spatial bandwidth h_j . A simple settlement-density component is

$$\pi_{\text{set}}(\mathbf{q}) = \sum_{j=1}^M (N_j + N_0)^\alpha \exp \left[-\frac{d_g(\mathbf{q}, \mathbf{c}_j)^2}{2h_j^2} \right], \quad (15)$$

where d_g is geographic distance. Gazetteers such as GeoNames or OpenStreetMap-derived settlement polygons can supply these coordinates [28], [29]. Population need not be used linearly; $0 < \alpha < 1$ prevents megacities from overwhelming all other evidence.

A node-type-conditioned prior can combine settlement, network-facility, transport/corridor, and land/water terms:

$$\pi_{\text{geo}}(\mathbf{q} | z_v) \propto w_{\text{land}}(\mathbf{q}) \left[\epsilon_z + a_z \pi_{\text{set}}(\mathbf{q}) + b_z \pi_{\text{net}}(\mathbf{q}) + c_z \pi_{\text{corr}}(\mathbf{q}) \right], \quad (16)$$

where z_v may distinguish access, aggregation, backbone, datacenter, or unknown node classes. The floor $\epsilon_z > 0$ prevents impossible certainty. Water receives a small rather than universally zero weight because submarine and coastal infrastructure exists. If node class is unknown, it is marginalized:

$$\pi_v(\mathbf{q}) = \sum_z p(z_v = z) \pi_{\text{geo}}(\mathbf{q} | z). \quad (17)$$

The target may use a different mixture from a backbone router. A residential endpoint prior may follow population more strongly, while a core-router prior may emphasize PoPs, IXPs, datacenters, long-haul corridors, and landing facilities. All spatial priors are normalized with respect to the chosen surface-area measure; in latitude/longitude coordinates

this must account for Earth geometry rather than treating degrees as equal-area Cartesian units. The hyperparameters $(a_z, b_z, c_z, \alpha, h_j)$ should be estimated or sensitivity-tested on held-out data. An over-strong urban prior can otherwise produce a precise but systematically biased posterior for rural or corridor infrastructure.

For an otherwise unanchored router,

$$p_0(\mathbf{q}_r) \propto \pi_r(\mathbf{q}_r), \quad (18)$$

and for the target

$$p_0(\mathbf{q}_X) \propto \pi_X(\mathbf{q}_X). \quad (19)$$

Fig. 2 illustrates the operation. The annulus is not deleted outside settlements; it is reweighted so that plausible inhabited/infrastructure locations receive more posterior mass while exceptional rural locations remain possible. Octant already demonstrated population and ocean constraints [12], and Spotter explicitly proposed population/geographic information as a prior [16]; Galeo's role for this term is therefore integration into every continuous latent-node factor rather than a novelty claim for demographic weighting itself.

V. CALIBRATION FROM THE PROBE MESH

A. Directed Probe-Pair Calibration

With n probes, a full directed mesh offers up to $n(n-1)$ ordered probe-pair measurement opportunities. For example, $n = 24$ yields 552 ordered pairs and $n = 36$ yields 1260 before repetition. These counts are not counts of independent constraints: many paths can merge onto the same backbone and therefore share latent segments and noise sources. Repeated OWD and traceroute observations nevertheless provide a dense calibration corpus

$$\mathcal{C} = \{Z_{ij}^{(r)}, \Pi_{i \rightarrow j}^{(r)}, O_{ij}^{(r)}\}, \quad (20)$$

where O denotes route-state/topology observations and external router-anchor evidence is collected separately as A .

For a stable observed path $\Pi = (e_1, \dots, e_m)$,

$$Z_{ij}^{(r)} = \sum_{e \in \Pi} \left(b_e + \beta_e d_e(Q) + q_e^{(r)} \right) + \epsilon_{ij}^{(r)}. \quad (21)$$

When route state is effectively observed, a calibration posterior can be written

$$p(\Theta, Q_{\mathcal{R}} \mid \mathcal{C}, A) \propto p(\Theta) p_0(Q_{\mathcal{R}}) p(A \mid Q_{\mathcal{R}}) \times \prod_{c \in \mathcal{C}} p(Z_c \mid \Pi_c, Q_{\mathcal{R}}, \Theta), \quad (22)$$

where Θ contains the hierarchical delay-distance coefficients, fixed offsets, queueing/noise parameters, clock terms, and route-selection parameters. If a calibration trace belongs to several plausible path families, the corresponding route state H_c is retained and summed/integrated rather than forcing one Π_c . The important point for later target inference is that (22) is a *joint* posterior: uncertainty in router geometry and delay parameters remains coupled.

Fig. 3 illustrates a directed probe-pair calibration relation after topology information has been incorporated. The synthetic data deliberately include positive delay outliers. Operational fitting should use the non-negative queueing model or a robust likelihood rather than interpreting an ordinary least-squares line as a propagation law.

B. Repeated Measurements and Uncertainty

For a stable quantity Y observed R times, define

$$\bar{Y} = \frac{1}{R} \sum_{r=1}^R Y^{(r)}, \quad s_Y^2 = \frac{1}{R-1} \sum_{r=1}^R (Y^{(r)} - \bar{Y})^2. \quad (23)$$

The sample variance characterizes network variability; the uncertainty of an estimated mean scales approximately as $s_Y/\sqrt{R_{\text{eff}}}$ only when an effective sample size R_{eff} accounts for temporal correlation and route-state persistence. Route changes are better modeled as mixtures or state transitions than as iid noise. A narrow posterior should therefore require both a stable measurement regime and enough effectively independent samples.

VI. SINGLE-SEGMENT AND PATH LIKELIHOODS

A. Generative Likelihood and Annular Potentials

The statistically clean starting point is a generative measurement likelihood. For a calibrated segment with range $r_e = d_e(Q)$ and observation y_e ,

$$p(y_e \mid \mathbf{q}_u, \mathbf{q}_v, \theta_e) = f_e(y_e - m_e(r_e; \theta_e)), \quad (24)$$

where m_e may use (10) and f_e contains the residual queueing, responder, and measurement model. Substituting $r_e = \|\mathbf{q}_u - \mathbf{q}_v\|$ makes (24) an annular likelihood in the unknown coordinate without requiring any additional Jacobian: it is a likelihood in the observation conditioned on position.

For visualization or for a calibrated inverse model, suppose the radial quantity itself is summarized as $\hat{d}_e$ with standard deviation σ_e . A convenient annular potential is

$$\psi_e(\mathbf{q}_u, \mathbf{q}_v) = \exp \left[-\frac{(\|\mathbf{q}_u - \mathbf{q}_v\|_2 - \hat{d}_e)^2}{2\sigma_e^2} \right]. \quad (25)$$

This is a Gaussian in the *radial residual*, not a bivariate Gaussian distribution. If instead one starts from a normalized radial posterior density $f_R(r \mid y)$ and wishes to interpret it as a two-dimensional spatial density with uniform bearing in a planar model, then with respect to area measure

$$p(\mathbf{q}_v \mid \mathbf{q}_u, y) = \frac{f_R(r \mid y)}{2\pi r}, \quad r = \|\mathbf{q}_v - \mathbf{q}_u\| > 0. \quad (26)$$

The circumference term must not be omitted in that inverse-density interpretation. Galeo therefore prefers the generative likelihood (24); empirical $p(r \mid y)$ models such as the Spotter-style relation are used as potentials only with their calibration prior and area measure made explicit.

For Internet measurements, a heavy-tailed radial residual is often safer. A Student- t potential is

$$\psi_e \propto \left(1 + \frac{(r_e - \hat{d}_e)^2}{\nu \sigma_e^2} \right)^{-(\nu+1)/2}, \quad (27)$$

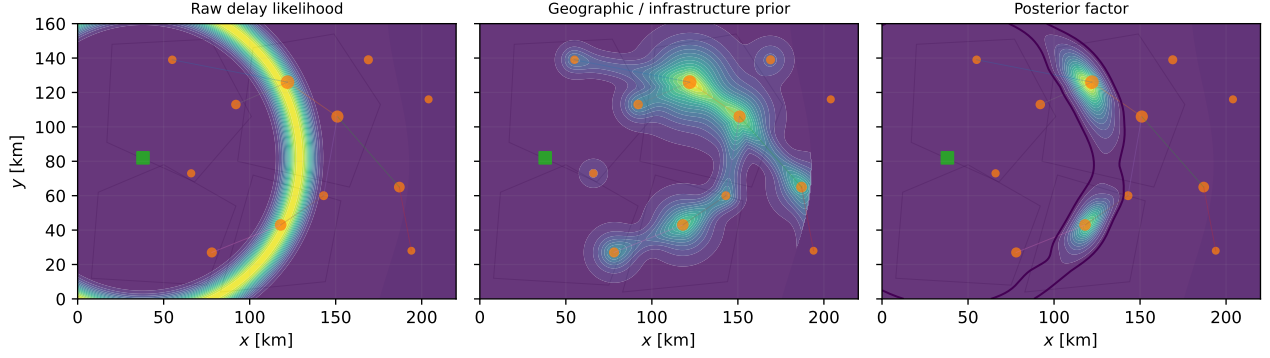


Fig. 2. Synthetic illustration of the Galeo geographic prior. Left: a raw Gaussian annular likelihood from timing evidence. Center: a soft prior built from settlement kernels, schematic network corridors, and a strongly down-weighted water region. Right: the product, normalized as a spatial factor. Probability is suppressed rather than forced to zero outside settlements.

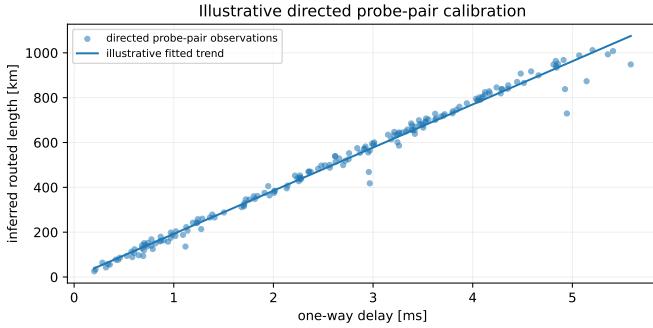


Fig. 3. Synthetic calibration illustration. Synchronized probe pairs provide one-way delay observations; routed length is constrained by topology and anchored geometry. The relation is statistical and pooled across path classes, not a universal propagation constant.

which reduces the influence of transient outliers or model mismatch.

B. Marginalizing Intermediate Routers

For a path from probe S_i to the target,

$$S_i = R_{i0} \rightarrow R_{i1} \rightarrow \dots \rightarrow R_{im_i} \rightarrow X, \quad (28)$$

let $Q_i = \{\mathbf{q}_{R_{i1}}, \dots, \mathbf{q}_{R_{im_i}}\}$. If this path forms a conditionally independent branch of the factor graph (or is being eliminated as a valid message conditioned on its separator variables), a target-side message has the form

$$m_{i \rightarrow X}(\mathbf{q}_X) = \int p(Q_i | \mathcal{C}, A) \prod_{e \in \Pi_i} \psi_e(Q_i, \mathbf{q}_X) dQ_i. \quad (29)$$

If two traceroutes share an uncertain router, independently computing (29) for both and then multiplying the results would generally double-count the shared latent uncertainty. Shared routers must be eliminated only once in the global factor graph (or through a valid junction/tree decomposition). The unified posterior in Sec. IX is therefore the normative definition;

per-path annuli are messages or visual summaries, not an independence assumption.

C. Discovered Routers as Probabilistic Secondary Landmarks

A central operational idea of Galeo is that probes do not only measure the final target X . They also repeatedly measure individual routers discovered on the intervening traceroute paths whenever those router addresses can themselves be targeted. A discovered router therefore moves from being a passive explanatory variable on one path to becoming an actively re-measured latent node shared by many paths.

Let R be such a discovered router and let $\mathcal{D}_R$ collect all measurements directed to R from the probe mesh together with path observations in which R appears. After conditioning on the calibration corpus $\mathcal{C}$ and anchor evidence A , Galeo forms a router posterior

$$p(\mathbf{q}_R | \mathcal{D}_R, \mathcal{C}, A) \propto \pi_R(\mathbf{q}_R) \prod_{j=1}^n m_{j \rightarrow R}(\mathbf{q}_R), \quad (30)$$

where $m_{j \rightarrow R}$ is the message contributed by probe S_j . Each message may itself integrate over upstream routers and route states; the important point is that all evidence about the same router is fused into one spatial posterior.

Once inferred, $p(\mathbf{q}_R | \mathcal{D}_R, \mathcal{C}, A)$ acts as a *soft anchor* or *probabilistic secondary landmark* for downstream localization. If R is adjacent or topologically close to the target, then instead of relying only on broad long-range direct-to-target evidence, Galeo can insert the short-hop factor from R to X :

$$p(\mathbf{q}_X | \mathcal{D}, \mathcal{C}, A) \propto \prod_{i \in I_{\text{dir}}} m_{i \rightarrow X}(\mathbf{q}_X) \times \int p(\mathbf{q}_R | \mathcal{D}_R, \mathcal{C}, A) \psi_{R,X}(\mathbf{q}_R, \mathbf{q}_X) d\mathbf{q}_R. \quad (31)$$

With several discovered routers near the target, the corresponding factors multiply. In the limiting case of a precisely known

router, $p(\mathbf{q}_R) = \delta(\mathbf{q}_R - \mathbf{q}_R^*)$, (31) reduces to an ordinary anchored hop factor.

This mechanism is the main reason Galeo can sharpen a target more than a purely direct long-range approach. A distant probe typically yields a broad annular likelihood for X ; a shorter downstream segment from a well-localized intermediate router yields a much steeper factor. Octant and CRG already demonstrated the value of secondary landmarks [12], [21]; Galeo's distinctive formulation is that the discovered router remains a continuous latent variable with a full posterior, shared globally across all paths, rather than being converted only to a point estimate or a coarse region.

VII. REGIME I: LOCAL NETWORK WITH VERIFIED SYMMETRY

Assume a regional or administratively controlled environment in which repeated measurements establish that forward and reverse paths use the same logical segments and that the two directions of each segment have approximately equal baseline data-plane delay. Under this condition,

$$R_{k+1 \rightarrow i} - R_{k \rightarrow i} \approx d_{k+1,k}^-, \quad (32)$$

and therefore from (3)

$$\begin{aligned} T_{i,k+1}^{\text{RTT}} - T_{ik}^{\text{RTT}} \\ \approx d_{k,k+1}^+ + d_{k+1,k}^- + (c_{i,k+1} - c_{ik}) + \Delta\eta. \end{aligned} \quad (33)$$

Only when responder-delay differences are negligible or removed by a stable low-quantile/robust procedure, and when $d_{k,k+1}^+ \approx d_{k+1,k}^-$, does the familiar approximation

$$\hat{\delta}_{k,k+1}^{\text{base}} \approx \frac{T_{i,k+1}^{\text{RTT}} - T_{ik}^{\text{RTT}}}{2} \quad (34)$$

represent a one-way baseline segment delay. It is an estimator under verified conditions, not a direct traceroute measurement. Persistent negative or inconsistent increments are diagnostics that these conditions do not hold.

The symmetry assumption must be tested rather than presumed. Probe-to-probe traceroutes in both directions should expose compatible router sequences after alias resolution, modulo non-responding interfaces and hidden segments. Repeated RTT distributions should also be stable enough that router response behavior does not dominate the increments. When these checks are satisfied, reverse-path state can be reduced or removed from the local model.

A. Local Synthetic Final-Stage Example

To illustrate the geometry without presenting synthetic data as empirical validation, consider a 120×100 km regional domain with 24 probes. The accompanying reproducible figure generator (fixed random seed 42) assumes that the full calibration/router-inference stage has already produced 24 approximately independent target-side radial messages. Their pseudo-ranges are generated from the true probe-to-target ranges plus $\mathcal{N}(0, 1.6^2)$ km perturbations, with radial standard deviations approximately 2.4–3.2 km. This is a *final-stage*

geometric illustration; it is not a packet-level simulation of the complete Galeo graph.

Four of the resulting messages are shown in Fig. 5. For this deliberately independent construction only, their normalized product is

$$p_{\text{ill}}(\mathbf{q}_X) = \frac{p_0(\mathbf{q}_X) \prod_i m_i(\mathbf{q}_X)}{\int_{\Omega} p_0(\mathbf{q}) \prod_i m_i(\mathbf{q}) d\mathbf{q}}. \quad (35)$$

The full Galeo model must instead use the joint marginalization in Sec. IX whenever messages share uncertain routers or calibration parameters.

In the realization in Fig. 6, the simulated target is (66, 44) km and the MAP of (35) is approximately (65.85, 44.15) km. The 95% HPD credible region occupies approximately 12.2 km². These values describe only the reproducible illustrative construction and must not be interpreted as an accuracy estimate for Galeo.

VIII. REGIME II: WIDE-AREA INTERNET WITH ROUTE ASYMMETRY

At Internet scale, the reverse path of an ICMP response cannot generally be assumed to be the inverse of the forward path. Large-scale reverse traceroute measurements demonstrate substantial path asymmetry [25]. The general model therefore keeps return-path contribution uncertain rather than deriving a forward link delay from adjacent RTTs.

For a traceroute observation to forward hop k , let H_{ik}^F be the observed or hypothesized forward prefix and let H_{ki}^R be a supported reverse-path hypothesis. The RTT likelihood is

$$\begin{aligned} p(T_{ik}^{\text{RTT}} | Q, \Theta, H_{ik}^F, H_{ki}^R) \\ = f_{\epsilon}(T_{ik}^{\text{RTT}} - \tau(H_{ik}^F) - \tau(H_{ki}^R) - c_{ik}), \end{aligned} \quad (36)$$

where c_{ik} represents responder/control-plane delay and f_{ϵ} is a Gaussian or robust residual model.

When repeated measurements, paired probes, reverse-traceroute capability, or other evidence identify a finite set of plausible return-path families, the reverse state can be marginalized:

$$\begin{aligned} p(T_{ik}^{\text{RTT}} | Q, \Theta, H_{ik}^F) \\ = \sum_{h \in \mathcal{H}_{ki}^R} p(h | \mathcal{C}) p(T_{ik}^{\text{RTT}} | Q, \Theta, H_{ik}^F, h). \end{aligned} \quad (37)$$

If the reverse topology is not observable, Galeo need not enumerate hypothetical Internet paths. The same uncertainty can be represented by an aggregate calibrated return-delay nuisance density $g_{ki}(r | \mathcal{C})$ and integrated over r . Likewise, if ECMP or route instability leaves the forward path uncertain,

$$p(T | Q, \Theta) = \sum_{f \in \mathcal{H}^F} \sum_{r \in \mathcal{H}^R} p(f, r | \mathcal{C}) p(T | Q, \Theta, f, r), \quad (38)$$

with the sums understood as path-family states supported by measurement, not an enumeration of all Internet routes. This construction converts routing ambiguity from an unmodeled error into posterior uncertainty.

Synthetic Galeo illustration: discovered intermediate routers sharpen the final target posterior

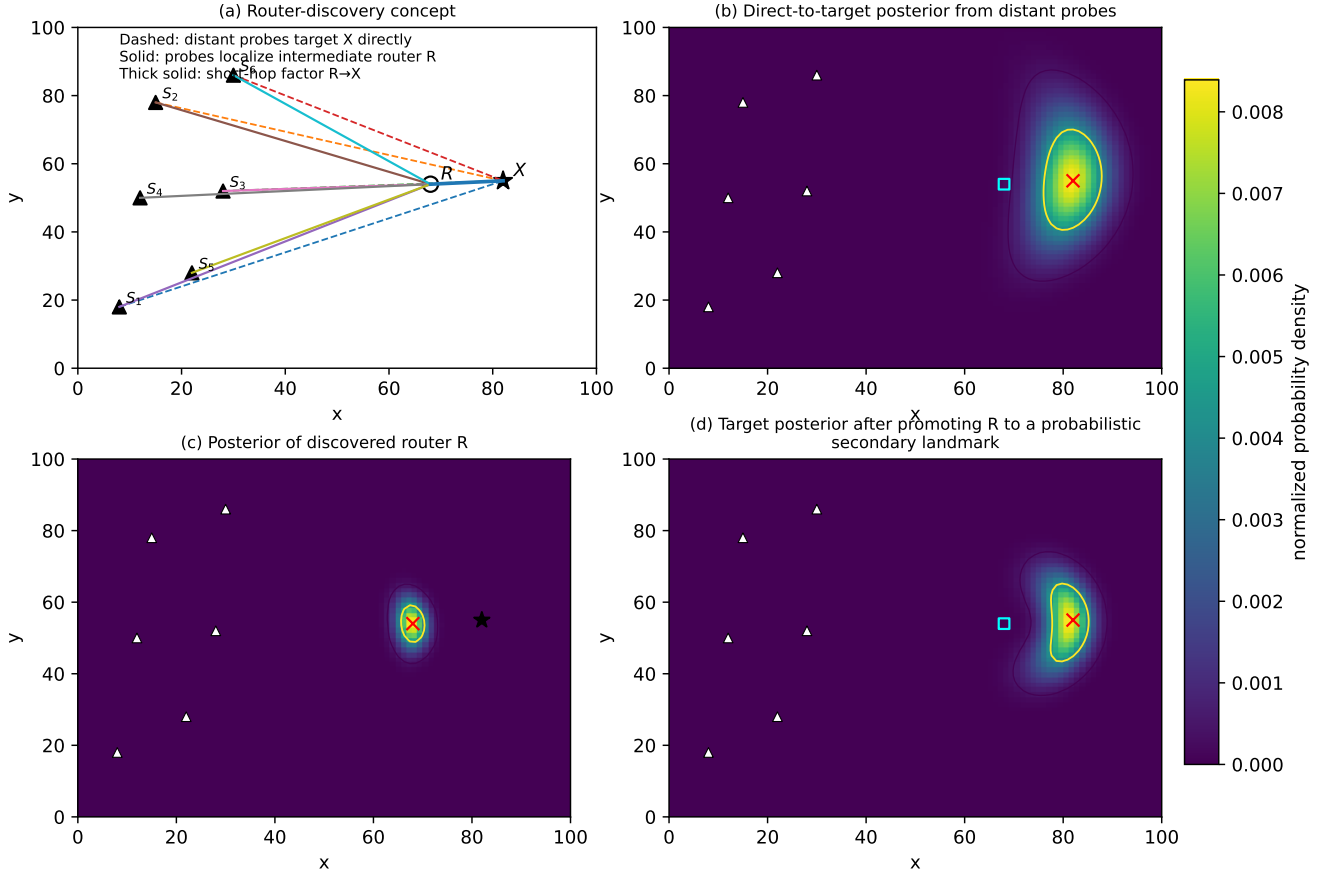


Fig. 4. Synthetic illustration of the Galeo secondary-landmark mechanism. (a) Several probes first discover an intermediate router R on paths toward the target X . Dashed lines denote direct long-range probe-to- X measurements; solid lines denote additional measurements aimed at R ; the thick segment is the short-hop factor from R to X . (b) Posterior for X when only the direct long-range probe-to-target evidence is used. (c) Posterior of the discovered router R after it is re-measured from multiple probes. (d) Posterior of X after promoting R to a probabilistic secondary landmark and combining $p(\mathbf{q}_R | \mathcal{D}_R)$ with the short-hop factor $\psi_{R,X}$. The target posterior becomes visibly sharper.

A. Learning a Local Symmetry Probability

A global Internet-wide symmetry constant is unnecessary. Where paired directional path evidence exists, define for an observed directed adjacency $u \rightarrow v$

$$\rho_{uv} = P(v \rightarrow u \text{ on reverse path} \mid u \rightarrow v \text{ on forward path}). \quad (39)$$

With a Beta prior $\rho_{uv} \sim \text{Beta}(\alpha, \beta)$, if K of N paired observations support the reverse adjacency,

$$\rho_{uv} \mid \mathcal{C} \sim \text{Beta}(\alpha + K, \beta + N - K). \quad (40)$$

This parameter is learned from paired path evidence, not inferred from a single RTT. It can identify regions where the local symmetric approximation is empirically defensible and regions where return-path uncertainty must remain explicit.

B. Wide-Area Synthetic Final-Stage Example

For a geometric illustration, consider a synthetic 1100×800 km region with 36 probes. As in the local example,

the figure generator assumes that router-level calibration and route-state elimination have already produced target-side radial messages. Each synthetic message is a two-component mixture representing a dominant path family and a broader displaced alternative:

$$m_i(d) = \rho_i \mathcal{N}(d; \mu_i, \sigma_{i,1}^2) + (1 - \rho_i) \mathcal{N}(d; \mu_i + \Delta_i, \sigma_{i,2}^2). \quad (41)$$

The fixed-seed source generates ρ_i , displacement, and scale parameters from the documented ranges in the accompanying script. The factors are deliberately final-stage messages rather than a synthetic packet-level Internet.

Fig. 7 shows four messages. Their secondary probability bands illustrate how route ambiguity can preserve alternative spatial explanations even when one path family dominates.

The illustrative normalized product is shown in Fig. 8. The simulated target is (655, 365) km; the MAP lies close to it in this particular fixed-seed realization, while the 95% HPD credible region remains much broader than in the local example. These quantities are demonstrations of posterior geometry

Selected synthetic final-stage radial messages

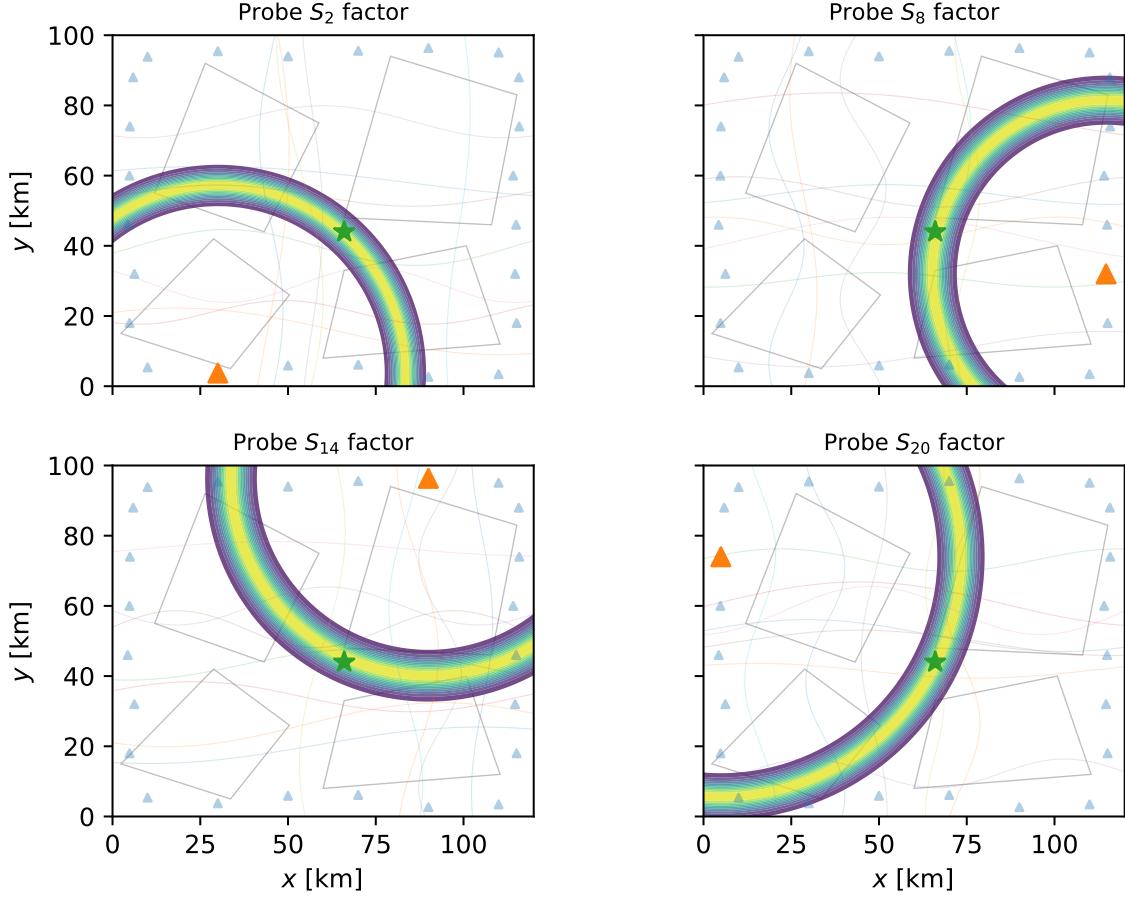


Fig. 5. Local symmetric final-stage illustration: four of 24 synthetic target-side radial messages. They represent the type of annular message that can emerge after router/calibration variables have been eliminated; they are not outputs of a packet-level network simulation.

only. Operational evaluation must report held-out credible-region coverage and point-error distributions rather than infer performance from this realization.

IX. UNIFIED BAYESIAN FORMULATION

Let $Q_R = \{\mathbf{q}_r : r \in \mathcal{R}\}$ denote all unknown router coordinates, let Z_R collect optional node-type variables for geographic priors, let H_C and H_D denote calibration and target route-family states, and let Θ contain hierarchical delay-distance, queueing, responder, clock, noise, alias-confidence, and route-selection parameters. Let A denote external router-anchor evidence. The statistically consistent definition uses

calibration data $\mathcal{C}$ and target data $\mathcal{D}$ in one joint posterior:

$$\begin{aligned}
 & p(\mathbf{q}_X, Q_R, Z_R, \Theta, H_C, H_D \mid \mathcal{C}, \mathcal{D}, A) \\
 & \propto p_0(\mathbf{q}_X) \prod_{r \in \mathcal{R}} p_0(\mathbf{q}_r \mid z_r) p(z_r) p(\Theta) \\
 & \quad \times p(A \mid Q_R) p(H_C, H_D \mid \Theta) \\
 & \quad \times \prod_{c \in \mathcal{C}} p(c \mid Q_R, \Theta, H_c) \\
 & \quad \times \prod_{m \in \mathcal{D}} p(m \mid \mathbf{q}_X, Q_R, \Theta, H_m). \tag{42}
 \end{aligned}$$

Here $p_0(\mathbf{q}_r \mid z_r)$ includes the soft geographic/infrastructure prior of (16), while $p(A \mid Q_R)$ contains precise or uncertain anchor factors such as (14). Interface-to-router alias states can be added to H or Θ when alias resolution is uncertain.

Equation (42) corrects an important bookkeeping issue: calibration does not yield independent marginals $p(\Theta \mid \mathcal{C})$ and $p(Q_R \mid \mathcal{C})$ that can safely be multiplied later. It yields a

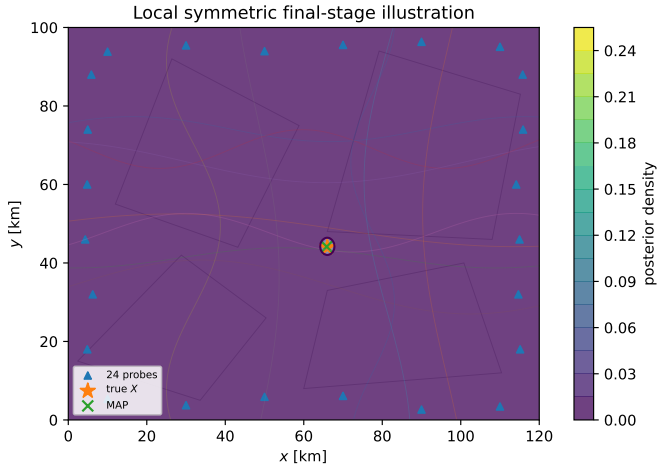


Fig. 6. Local symmetric final-stage illustration with 24 probes. Filled contours show the normalized illustrative target posterior; the outer contour marks the 95% HPD credible set. The star is the simulated target and the cross is the MAP estimate. Background lines are schematic, not real geography.

coupled posterior. An equivalent staged implementation first computes

$$p(Q_R, Z_R, \Theta, H_C | \mathcal{C}, A) \quad (43)$$

and then carries this *joint* distribution forward as the prior state for target inference.

Let $\Xi = (Q_R, Z_R, \Theta, H_C, H_D)$ collect all nuisance variables and let $d\mu(\Xi)$ denote integration over continuous components and summation over discrete components. The desired target geolocation density is obtained by eliminating every nuisance variable exactly once:

$$p_X(x, y | \mathcal{D}, \mathcal{C}, A) = \int p(\mathbf{q}_X, \Xi | \mathcal{C}, \mathcal{D}, A) d\mu(\Xi), \quad (44)$$

with $\mathbf{q}_X = (x, y)$. Equation (44) is the core definition of Galeo.

For compact notation, let $\Xi_C = (Q_R, Z_R, \Theta, H_C)$ denote the calibrated latent state. Define the *global* target evidence

$$\Phi(\mathbf{q}_X) = \int p(\Xi_C | \mathcal{C}, A) \left[\sum_{H_D} p(H_D | \Xi_C) \times \prod_{m \in \mathcal{D}} p(m | \mathbf{q}_X, \Xi_C, H_m) \right] d\mu(\Xi_C). \quad (45)$$

Then

$$p_X(\mathbf{q}_X | \mathcal{D}, \mathcal{C}, A) = \frac{p_0(\mathbf{q}_X) \Phi(\mathbf{q}_X)}{\int_{\Omega} p_0(\mathbf{q}) \Phi(\mathbf{q}) d\mathbf{q}} \quad (46)$$

without assuming that separately marginalized probe paths are independent.

The MAP point is

$$\hat{\mathbf{q}}_X^{\text{MAP}} = \arg \max_{\mathbf{q}_X} p_X(\mathbf{q}_X | \mathcal{D}, \mathcal{C}, A). \quad (47)$$

For posterior uncertainty reporting, define the $100\gamma\%$ HPD credible set

$$C_\gamma = \{\mathbf{q} : p_X(\mathbf{q} | \mathcal{D}, \mathcal{C}, A) \geq c_\gamma\}, \quad (48)$$

where c_γ satisfies

$$\int_{C_\gamma} p_X(\mathbf{q} | \mathcal{D}, \mathcal{C}, A) d\mathbf{q} = \gamma. \quad (49)$$

X. SEQUENTIAL INTERPRETATION

The original annulus-propagation intuition can be written as a message-passing recursion. If $f_k(\mathbf{q})$ is the spatial density of hop k , then on a chain

$$f_{k+1}(\mathbf{q}') = \frac{1}{Z_k} \int_{\Omega} f_k(\mathbf{q}) \psi_k(\mathbf{q}, \mathbf{q}') d\mathbf{q}, \quad (50)$$

starting from

$$f_0(\mathbf{q}) = \delta(\mathbf{q} - \mathbf{q}_{S_i}). \quad (51)$$

Equation (50) is an exact continuous-space sum-product step for a chain (and, with the usual separator bookkeeping, for a tree/junction tree). Intersecting traceroutes create loops. Ordinary loopy belief propagation on such a graph is then approximate, so convergence and calibration must be checked; MCMC, sequential Monte Carlo, variational methods, or junction-tree/elimination methods remain alternatives. This distinction prevents “message passing” from being mistaken for an exact algorithm on an arbitrary Galeo graph.

XI. INFERENCE STRATEGY

Exact evaluation of (44) is generally infeasible because the number of routers, alias hypotheses, and route states can be large. A practical implementation can combine the following methods.

- 1) **Topology normalization.** Collect repeated Paris-style traceroutes; resolve interface aliases where evidence is strong; retain uncertain alias groups probabilistically; and represent MPLS/non-responding sections as aggregate logical segments rather than invented physical links.
- 2) **Route-family construction.** Cluster repeated forward traces into supported path families. Where reverse-path evidence exists, construct corresponding return families; otherwise fit an aggregate return-delay nuisance model.
- 3) **Calibration stage.** Fit the joint state $p(Q_R, Z_R, \Theta, H_C | \mathcal{C}, A)$ using pooled/hierarchical delay-distance parameters. Minimum or low-quantile observations can inform baseline delay, while positive queueing and control-plane residuals are modeled separately.
- 4) **Graph reduction.** Eliminate well-calibrated internal subgraphs into separator-preserving messages while retaining common routers that couple measurement paths. Reduction must not independently marginalize the same shared router into several factors.
- 5) **Target inference.** Evaluate (44) with particle methods, MCMC, sequential Monte Carlo, variational inference, discretized message passing, junction-tree methods where feasible, or a hybrid scheme.
- 6) **Robustification.** Use a heavy-tailed/outlier component and route-state mixtures so that one pathological path, ICMP response, or transient queueing event cannot collapse the posterior.

Wide-area synthetic final-stage mixture messages

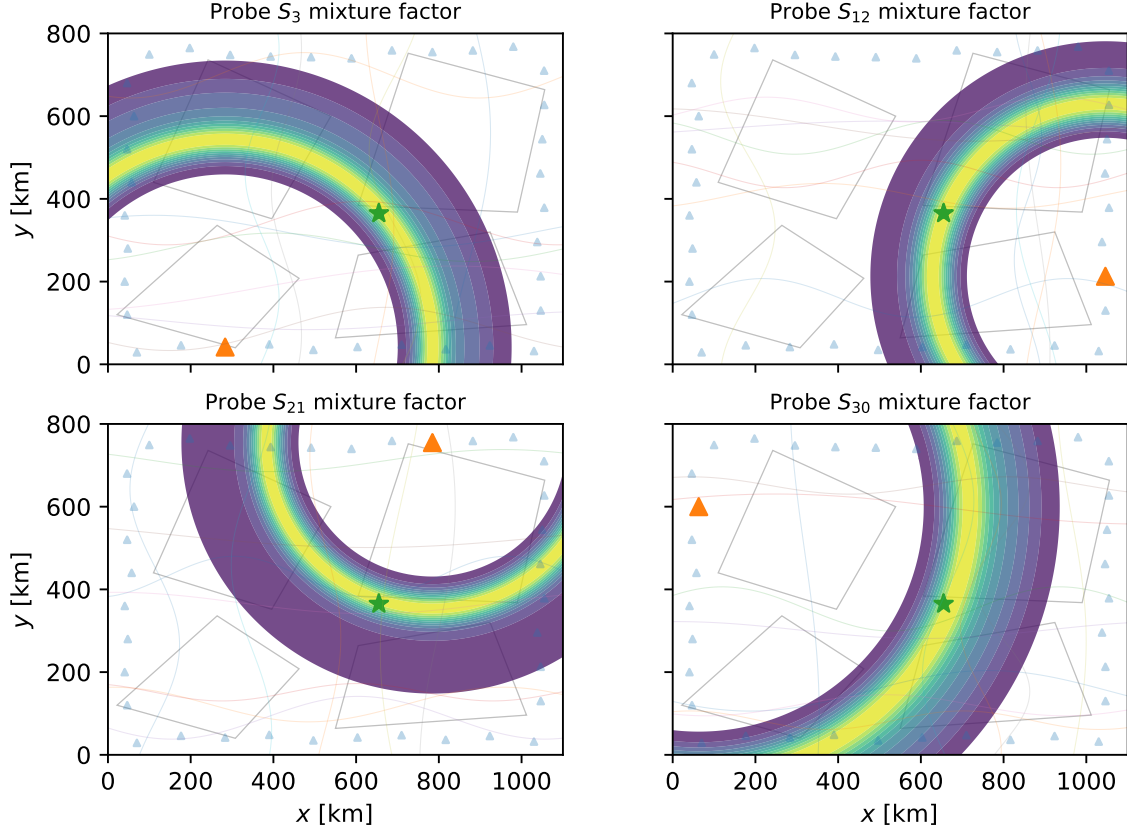


Fig. 7. Wide-area final-stage illustration: four synthetic mixture messages representing uncertainty after path/return-path marginalization. The wider or displaced bands are route-state uncertainty, not Gaussian measurement noise alone.

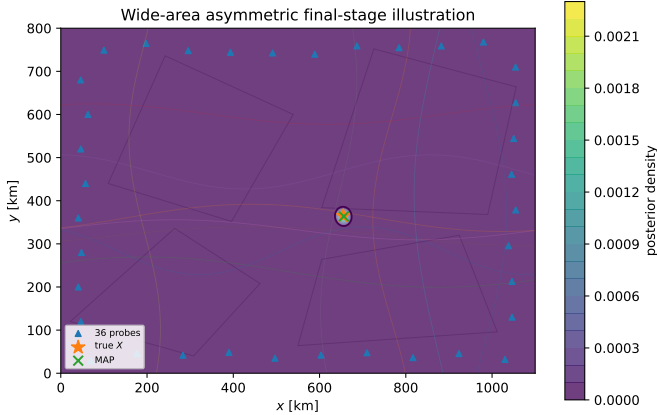


Fig. 8. Wide-area asymmetric final-stage illustration using 36 probes. Route-family mixtures produce a visibly broader 95% HPD credible region than the verified-symmetry example. Background lines are schematic, not real geography.

- 7) **Reporting and diagnostics.** Return MAP/mean coordinates, HPD credible regions, posterior entropy, effective

independent path groups, alias/route-state uncertainty, and diagnostics for model mismatch and posterior predictive calibration.

For a two-dimensional target with tens of probes, evaluating a grid over X is inexpensive compared with integrating the shared router/calibration state. High-dimensional calibration can therefore be performed offline, but any compact online representation must preserve the dependence induced by shared routers (for example through separator messages or a joint surrogate), rather than exporting independently marginalized per-probe rings and multiplying them naively.

XII. PROBE COUNT AND PLACEMENT

Galeo was conceived for a serious distributed apparatus with at least lower tens of probes, and the illustrative examples use 24 and 36. These numbers are design examples, not a universal statistical minimum. Required probe count depends on angular geometry, topology overlap, anchor density, route stability, timing quality, and the desired credible-region coverage.

For n probes, a full directed schedule contains

$$N_{\text{pair}} = n(n - 1) \quad (52)$$

ordered probe-pair opportunities. Thus 20, 24, 32, and 36 probes yield 380, 552, 992, and 1260 ordered pairs, respectively. The raw count grows quadratically, but the *effective* information does not: many pairs may share the same provider, metropolitan egress, backbone, or route family. Measurement scheduling should therefore favor spatial and topological diversity rather than maximize pair count alone.

A target should ideally be constrained by probes from several azimuths and several independent network ingress directions. Useful deployment diagnostics include angular coverage around the target hypothesis, pairwise baseline diversity, AS/provider diversity, number of distinct route families/common-router separators, and an effective number of independent path groups. Adding ten probes behind one provider and one egress may contribute less information than a few probes with genuinely different geography and topology.

XIII. IDENTIFIABILITY AND FAILURE MODES

The model makes uncertainty explicit but does not eliminate fundamental ambiguity. Important limitations are:

A. Delay–Geometry Confounding

Two geographically different paths can have similar delay, and a longer physical route can be faster than a shorter congested route. More fundamentally, if every logical segment were assigned an unconstrained private β_e , the product $\beta_e d_e$ in (10) could fit many geometries. Probe coordinates fix the global spatial reference frame, but intermediate-router geometry is identifiable only where overlapping path equations, anchors, and the hierarchical calibration of (12) provide sufficient rank/observability. The posterior can remain broad or multimodal; a correct implementation must preserve that ambiguity rather than force a single Gaussian location.

B. Hidden Topology and Logical Segments

Traceroute observes IP-layer responders, not every physical forwarding element. MPLS tunnels, non-responding routers, layer-2 transport, and ICMP generation behavior can hide or distort the path graph. Such sections should remain aggregate latent segments unless additional measurements reveal their contents.

C. Router Interface Aliasing

Different IP interfaces can belong to one physical router, and a router can source ICMP from an interface different from the one receiving the probe. Treating interface addresses as distinct physical routers creates false geometric constraints. Alias resolution must carry confidence; uncertain mappings should remain latent rather than being converted into hard topology.

D. Target Semantics, Anycast, NAT, Tunnels, and Proxies

Galeo estimates the location of the network-visible responding IP endpoint/instance described by its measurement model. That point may differ from the physical user or application service behind NAT, a proxy, a tunnel, or a load balancer. Anycast requires conditioning on the responding instance/path

state, and a filtered target may leave only a last-responsive-hop constraint rather than a direct endpoint observation.

E. Clock Error

OWD accuracy is limited by synchronization. If source and destination clock offsets are c_i and c_j , then

$$Z_{ij}^{\text{obs}} = Z_{ij}^{\text{true}} + (c_j - c_i) + \epsilon. \quad (53)$$

Clock offset and drift should therefore be included in Θ unless bounded tightly by GNSS/PTP or another timing architecture.

F. Correlated Evidence

Traceroutes from different probes are not conditionally independent when they merge onto shared infrastructure. Naively multiplying independently marginalized path likelihoods becomes overconfident. Shared-router inference in (42) is specifically intended to capture this dependence; any approximation must preserve or explicitly temper the residual correlation.

G. Geographic-Prior Bias

Population and infrastructure priors can materially improve localization, but they can also dominate weak timing evidence. A prior trained on city-heavy data may under-cover rural cabinets, long-haul corridors, landing stations, or remote access networks. Prior weights and bandwidths therefore require held-out calibration and sensitivity analysis, and an explicit non-zero background component should remain.

XIV. EVALUATION METHODOLOGY

An empirical study should separate calibration, hyperparameter/prior tuning, and target evaluation sets. Ground-truth targets should span local, regional, national, and continental distances; multiple AS/provider paths; urban and rural infrastructure; and route-stable as well as route-changing conditions. Recommended metrics include:

- median, 75th, 90th, and 95th percentile point error;
- area of the 50%, 90%, and 95% HPD credible regions;
- empirical coverage of each nominal credible region;
- posterior predictive calibration of delay/RTT distributions;
- sensitivity to probe count, angular placement, and effective independent path groups;
- sensitivity to router-anchor density, alias uncertainty, and anchor error;
- sensitivity to settlement/infrastructure prior strength and node-type assumptions;
- performance under route changes, ECMP, asymmetric return paths, hidden MPLS segments, and ICMP outliers;
- ablations of OWD calibration, shared-router coupling, geographic priors, alias uncertainty, and return-path marginalization.

A method whose nominal 95% credible region contains only 60% of held-out ground-truth targets is overconfident even if its median point error is small. Credible-region calibration and posterior predictive checks should therefore be first-class results. The synthetic figures in this paper are reproducible

geometry demonstrations only and do not substitute for such validation.

XV. RETROSPECTIVE POSITIONING AND SCOPE OF THE ARCHIVE

The comparison in Table I changes how Galeo should be described. It would be inaccurate to claim novelty for topology-aware geolocation, intermediate-router localization, probabilistic delay-distance mapping, Gaussian annuli, demographic/geographic constraints, or factor-graph reasoning individually. All have verified precedents.

The archival value of Galeo lies in the specific integrated generative interpretation: a repeated traceroute corpus defines a persistent graph; each physical router is represented once; every unknown router carries a continuous spatial density; calibrated hop/path observations contribute probabilistic factors; intersecting routes exchange information through shared router variables; known routers enter as anchors; geography enters as a soft prior for every relevant node; and unsupported path symmetry is represented by latent route states rather than silently assumed. In the local symmetric case this reduces to the original recursive annulus intuition in (50); in the wide-area case the same model survives by marginalization over additional state.

The reviewed literature contains systems that cover large subsets of this design. TBG is closest on simultaneous intermediate-router topology [11]; Octant on transitive router localization plus geographic constraints [12]; the Laki path model on OWD-calibrated router geometry [14], [15]; Spotter on Gaussian annular probability surfaces [16]; Chandekar and Paris on joint factor-graph inference [17]; CRG on statistically localized common routers as secondary landmarks [21]; and mi^2 on belief propagation through localized Internet topology to discrete colocation facilities [22]. The absence of an exact full combination in this review should be interpreted conservatively. This paper therefore records Galeo as an earlier unpublished method in a form that can be reproduced and compared, without asserting legal novelty or priority over independent publications.

XVI. CONCLUSION

Galeo is archived here as a topology-aware Bayesian geolocation method conceived in earlier unpublished work and formalized for publication in 2026. Its central proposition is that a distributed measurement system should not discard the router topology exposed by traceroute. Known probes and router anchors fix portions of the graph in space; repeated probe-to-probe measurements calibrate directed delay behavior; every shared intermediate router remains one latent continuous spatial variable; and target localization follows by propagating and marginalizing uncertainty over the coupled graph.

The method naturally yields annular likelihoods or radial potentials rather than deterministic radii. In a verified symmetric local network, successive traceroute RTTs can support an approximate incremental-link estimator only after alias resolution, path-symmetry checks, and control-plane-delay filtering. In the general Internet, reverse paths, ECMP/route changes, hidden

topology, ICMP response behavior, and queueing prevent that reduction; route states and nuisance parameters must instead be integrated out, producing broader or multimodal posteriors. A settlement/infrastructure prior can substantially reweight otherwise geometrically equivalent annulus segments, but it must remain soft and externally calibrated so that rural and exceptional infrastructure is not made impossible.

Retrospective comparison shows that many Galeo ingredients were independently published in closely related work, including simultaneous router localization, transitive secondary landmarks, one-way path calibration, Gaussian delay-distance probability models, demographic constraints, and factor-graph inference. Accordingly, the contribution of this paper is documentary: it provides a precise, reproducible statement of Galeo as an integrated continuous generative model and records how that earlier unpublished conception relates to the verified literature. A future empirical implementation could test the formulation using a dense probe mesh, synchronized OWD where available, modern Paris-style traceroute, router alias resolution, and geographically grounded priors; such validation is deliberately outside the archival claim made here.

REFERENCES

- [1] S. Niccolini, S. Tartarelli, J. Quittek, T. Dietz, and M. Swany, "Information Model and XML Data Model for Traceroute Measurements," RFC 5388, Dec. 2008, doi: 10.17487/RFC5388.
- [2] F. Baker, Ed., "Requirements for IP Version 4 Routers," RFC 1812, Jun. 1995, doi: 10.17487/RFC1812.
- [3] G. Almes, S. Kalidindi, M. Zekauskas, and A. Morton, Ed., "A One-Way Delay Metric for IP Performance Metrics (IPPM)," RFC 7679, STD 81, Jan. 2016, doi: 10.17487/RFC7679.
- [4] S. Shalunov, B. Teitelbaum, A. Karp, J. Boote, and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)," RFC 4656, Sep. 2006, doi: 10.17487/RFC4656.
- [5] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding traceroute anomalies with Paris traceroute," in *Proc. ACM Internet Measurement Conf. (IMC)*, 2006, pp. 153–158, doi: 10.1145/1177080.1177100.
- [6] K. Keys, Y. Hyun, M. Luckie, and K. Claffy, "Internet-Scale IPv4 Alias Resolution With MIDAR," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 383–399, Apr. 2013, doi: 10.1109/TNET.2012.2198887.
- [7] Y. Vanaubel, J.-R. Luttinger, P. Mérindol, J.-J. Pansiot, and B. Donnet, "TNT, Watch Me Explode: A Light in the Dark for Revealing MPLS Tunnels," in *Proc. Network Traffic Measurement and Analysis Conf. (TMA)*, 2019, pp. 65–72, doi: 10.23919/TMA.2019.8784525.
- [8] V. N. Padmanabhan and L. Subramanian, "An investigation of geographic mapping techniques for Internet hosts," in *Proc. ACM SIGCOMM*, 2001, pp. 173–185, doi: 10.1145/383059.383073.
- [9] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-Based Geolocation of Internet Hosts," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1219–1232, Dec. 2006, doi: 10.1109/TNET.2006.886332.
- [10] B. Gueye, S. Uhlig, A. Ziviani, and S. Fdida, "Leveraging buffering delay estimation for geolocation of Internet hosts," in *NETWORKING 2006*, LNCS 3976, 2006, pp. 319–330, doi: 10.1007/11753810_27.
- [11] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, "Towards IP geolocation using delay and topology measurements," in *Proc. ACM Internet Measurement Conf. (IMC)*, 2006, pp. 71–84, doi: 10.1145/1177080.1177090.
- [12] B. Wong, I. Stoyanov, and E. G. Sirer, "Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts," in *Proc. 4th USENIX Symp. Networked Systems Design and Implementation (NSDI)*, 2007, pp. 313–326.
- [13] I. Youn, B. L. Mark, and D. Richards, "Statistical Geolocation of Internet Hosts," in *Proc. 18th Int. Conf. Computer Communications and Networks (ICCCN)*, 2009, pp. 1–6, doi: 10.1109/ICCCN.2009.5235373.

- [14] S. Laki, P. Mátray, P. Haga, I. Csabai, and G. Vattay, "A Detailed Path-Latency Model for Router Geolocation," in *Proc. 5th Int. Conf. Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCom)*, 2009, pp. 1–6, doi: 10.1109/TRIDENT-COM.2009.4976258.
- [15] S. Laki, P. Mátray, P. Haga, I. Csabai, and G. Vattay, "A model based approach for improving router geolocation," *Computer Networks*, vol. 54, no. 9, pp. 1490–1501, 2010, doi: 10.1016/j.comnet.2009.12.004.
- [16] S. Laki, P. Mátray, P. Haga, T. Sebok, I. Csabai, and G. Vattay, "Spotter: A Model Based Active Geolocation Service," in *Proc. IEEE INFOCOM*, 2011, pp. 3173–3181, doi: 10.1109/INFCOM.2011.5935165.
- [17] S. Chandekar and B.-P. Paris, "Large-scale, discrete IP geolocation via multi-factor evidence fusion using factor graphs," in *Proc. 18th Int. Conf. Information Fusion (FUSION)*, Washington, DC, USA, 2015, pp. 1497–1504.
- [18] K. B. Laskey, S. Chandekar, and B.-P. Paris, "A probabilistic ontology for large-scale IP geolocation," in *Proc. 10th Conf. Semantic Technologies for Intelligence, Defense, and Security (STIDS)*, CEUR-WS, vol. 1523, 2015, pp. 18–25.
- [19] J. Chen, F. Liu, X. Luo, F. Zhao, and G. Zhu, "A landmark calibration-based IP geolocation approach," *EURASIP Journal on Information Security*, vol. 2016, art. 4, 2016, doi: 10.1186/s13635-015-0029-5.
- [20] J.-N. Chen, F.-L. Liu, Y.-F. Shi, and X. Luo, "Towards IP Location Estimation Using the Nearest Common Router," *Journal of Internet Technology*, vol. 19, no. 7, pp. 2097–2110, 2018, doi: 10.3966/160792642018121907013.
- [21] Z. Wang, H. Li, Q. Li, H. Zhu, W. Li, and L. Sun, "Towards IP geolocation with intermediate routers based on topology discovery," *Cybersecurity*, vol. 2, art. 13, 2019, doi: 10.1186/s42400-019-0030-2.
- [22] R. Motamedi, B. Yeganeh, B. Chandrasekaran, R. Rejaie, B. M. Maggs, and W. Willinger, "On Mapping the Interconnections in Today's Internet," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 2056–2070, 2019, doi: 10.1109/TNET.2019.2940369.
- [23] Z. Wang, F. Zhou, W. Zeng, G. Trajcevski, C. Xiao, Y. Wang, and K. Chen, "Connecting the Hosts: Street-Level IP Geolocation with Graph Neural Networks," in *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, 2022, pp. 4121–4131, doi: 10.1145/3534678.3539049.
- [24] S. Ding, X. Luo, J. Wang, and X. Fu, "GNN-Geo: A Graph Neural Network-Based Fine-Grained IP Geolocation Framework," *IEEE Trans. Network Science and Engineering*, vol. 10, 2023, doi: 10.1109/TNSE.2023.3266752.
- [25] K. Vermeulen, E. Gurmericililer, I. Cunha, D. Choffnes, and E. Katz-Bassett, "Internet Scale Reverse Traceroute," in *Proc. ACM Internet Measurement Conf. (IMC)*, 2022, pp. 694–715, doi: 10.1145/3517745.3561422.
- [26] X. Liu, W. Tai, T. Zhong, Y. Wang, K. Chen, and F. Zhou, "Tracing paths, pruning noise: Toward robust IP geolocation via topology-guided shaping and refinement," *Information Processing & Management*, vol. 63, no. 3, art. 104504, 2026, doi: 10.1016/j.ipm.2025.104504.
- [27] A. Zilberman, A. Offer, B. Pincu, Y. Glickshtein, R. Kant, O. Brodt, A. Otung, R. Puzis, A. Shabtai, and Y. Elovici, "A Survey on Geolocation on the Internet," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 5, pp. 3339–3381, 2025, doi: 10.1109/COMST.2024.3518398.
- [28] GeoNames, "GeoNames geographical database," settlement and population gazetteer. [Online]. Available: <https://www.geonames.org/>. Accessed: Sep. 2026.
- [29] OpenStreetMap contributors, "OpenStreetMap," geographic and settlement data. [Online]. Available: <https://www.openstreetmap.org/>. Accessed: Sep. 2026.