NetBreakout: A Scalable Methodology for Mapping Mobile Network Egress Topology

Syed Tauhidun Nabi
Virginia Tech
tauhidun@vt.edu

Shaddi Hasan
Virginia Tech
shaddi@vt.edu

ABSTRACT

Despite their critical role in global connectivity, the internal topologies of mobile networks remain largely opaque. Packet Data Network Gateways (P-GWs/UPFs) act as the critical egress points where subscriber traffic exits to the Internet, yet little is known about their number, placement, or evolution over time. This lack of visibility limits independent assessments of resilience, latency, and equity in mobile service delivery.

We present *NetBreakout*, an ongoing effort to design a scalable, reproducible methodology for mapping mobile egress infrastructure. Our approach begins with inside-out ground truth from subscriber vantage points and generalizes outward using public datasets such as CAIDA's ITDK [2] and M-Lab's scamper traceroutes [11]. We also leverage the historical Mobiperf dataset (2012–2020) [13], which contains around 370,000 mobile traceroutes, as a validation baseline. By focusing on the P-GW as the stable anchor between mobile access and the Internet core, *NetBreakout* provides new, policy-relevant visibility into critical infrastructure. We seek community feedback on methodological design, validation strategies, and policy applications.

1 INTRODUCTION

Mobile networks are no longer a mere convenience; they are a layer of critical infrastructure underpinning the global economy, emergency services, and daily life. The COVID-19 pandemic accelerated this dependence, cementing mobile broadband (MBB) as an essential lifeline for healthcare, education, and finance, especially in regions lacking robust fixed-line alternatives [12, 14].

Although fixed broadband still carries the majority of internet traffic [4], the unique flexibility of MBB makes it indispensable for bridging the digital divide. In countries like Bangladesh, for example, MBB constitutes over 90% of Internet access, playing a critical role in socio-economic participation [1, 19]. However, this deep reliance exposes a critical vulnerability: a failure in the mobile core network can paralyze entire regions, as dramatically demonstrated when a single facility failure in Nashville, Tennessee in December 2020 disconnected 911 services, air traffic control, and hospital networks [10, 16]. Understanding the architecture of this vital yet opaque infrastructure is therefore a pressing concern.

This fundamental opacity presents a significant challenge for researchers, regulators, and the public. The specific topology of the mobile packet core—the infrastructure that bridges the Radio Access Network (RAN) with the global internet—is proprietary and not publicly disclosed. Central to this core is the *Packet Data Network Gateway (P-GW)*[1], which serves as the anchor point for user mobility and acts as the critical egress point, or "breakout," for all subscriber traffic destined for the internet. Without a clear map of this egress topology, it is difficult to independently assess network resilience against failures, analyze sources of end-to-end latency, or track how operators are evolving their infrastructure to support new demands like 5G and edge computing.

Research Questions. This work addresses three fundamental questions:

(1) **Identification:** Can we design a scalable methodology to infer the number and placement of P-GWs within a mobile operator using primarily public data?
(2) **Longitudinal tracking:** Can the methodology detect and validate changes in P-GW deployment over time, enabling independent monitoring of infrastructure growth and resilience?
(3) **Geolocation:** To what accuracy and at what granularity can we geolocate P-GWs?

By combining controlled *inside-out* active measurements from subscriber vantage points with scalable *outside-in* inference using public datasets, *NetBreakout* provides the first hybrid, mobile-centric view of Internet egress infrastructure. We argue that such visibility is vital not only for academic research, but also for strengthening resiliency planning and digital equity policy.

2 RELATED WORK

Early ISP Topology Mapping. The problem of mapping ISP infrastructure from the outside traces back to *Rocketfuel* [18], which combined BGP-informed *directed probing*, path reduction heuristics, and novel alias resolution techniques to construct router-level maps of ten large ISPs. Rocketfuel also pioneered the use of DNS hostnames to infer router roles and locations, releasing the first realistic ISP maps to the community. Despite its impact, Rocketfuel required heavy active probing and was not tailored to mobile networks, where subscriber traffic traverses opaque core architectures hidden behind CGNAT and NAT444 [15].

Router Ownership and AS Boundary Inference. Subsequent work focused on improving the accuracy of inferring interdomain borders. The *bdrmapIT* tool [8] combined heuristics from earlier efforts (bdrmap [5] and MAP-IT [9]) to infer router ownership and AS boundaries at Internet scale with 91.8–98.8% accuracy. By systematically identifying the last hop within an AS before traffic exits, bdrmapIT provides the state-of-the-art methodology for detecting border routers.

DNS-based Geolocation. Rocketfuel's *undns* database of manually-crafted regexes seeded two decades of work on hostname-based inference, but quickly became outdated. More recently in 2020, Luckie et al. introduced *Hoiho* [6, 7], which automatically learns regexes for extracting geohints from rDNS. Validated with operator

[1] In 5G networks, the analogous function is provided by the User Plane Function (UPF). In this paper, we use "P-GW" as a common term to refer to these egress gateways across both 4G and 5G contexts.

ground truth, Hoiho achieves 94% accuracy in interpreting router geolocation strings, outperforming earlier systems such as DRoP [3] and HLOC [17]. We adopt Hoiho's methodology to corroborate the location of P-GWs, while also noting its limitations in the presence of operator-specific naming deviations.

The Unique Challenges of Mobile Network Cartography. These state-of-the-art topology inference techniques face fundamental obstacles when applied to mobile networks because of their distinct architectural and operational properties. First, mobile cores are highly opaque: unlike fixed-line ISPs, they rarely use informative DNS conventions and often block external probes at their borders. Second, the subscriber edge is particularly difficult to measure. Carrier-grade NAT (CGNAT) is nearly ubiquitous—Richter et al. [15] found that over 90% of cellular ASes deploy CGN, often with multiple layers of NAT444—while operator policies that block ICMP responses prevent direct reachability testing of end users. This challenge was highlighted in *Access Denied* [10], where the authors had to rely on coarse-grained CAIDA Ark vantage points to study AT&T's network, since mobile subscribers—unlike cable users—did not respond to pings.

Finally, the scarcity of public "inside-out" vantage points forces researchers to develop creative but bespoke data collection campaigns. The IMC 2021 study [20] introduced the innovative ShipTraceroute experiment to gain nationwide visibility by shipping smartphones across the country. Similarly, both papers [10, 20] resorted to wardriving public WiFi hotspots to acquire the necessary local vantage points to map service areas. While these active measurement approaches have yielded valuable insights, they remain labor-intensive campaigns rather than scalable, repeatable frameworks. This underscores a central gap in the literature: a primarily passive methodology for systematically mapping mobile network egress topology does not yet exist.

Synthesis. In summary, *Rocketfuel* demonstrated router-level ISP mapping, *bdrmapIT* advanced AS-boundary inference, and *Hoiho* modernized DNS-based geolocation. Yet mobile-specific studies remain limited, as CGNAT and blocked probes obscure subscriber-facing infrastructure. To address this gap, *NetBreakout* aims to combine inside-out ground truth from subscriber vantage points with scalable outside-in inference using public datasets. This hybrid approach enables the identification, longitudinal tracking and geolocation of P-GWs—the critical egress anchors where mobile access meets the Internet core—offering policy-relevant visibility into infrastructure that is vital yet historically opaque.

3 METHODOLOGY

Our approach is guided by two realities: (i) mobile operators rarely expose their core topology to external vantage points, often blocking ICMP probes at their network borders, and (ii) purely "outside-in" techniques (e.g., analyzing traceroute data from M-Lab [11] or CAIDA's ITDK [2]) risk misclassifying firewalls or filtered interfaces as Packet Gateways. To address these limitations, we adopt an *inside-out first* methodology that begins with ground truth from subscriber vantage points within the mobile network, and then generalizes outward to large-scale public datasets for scalability.

Phase 1: Ground-truth seeding (inside-out)

We deploy controlled subscriber vantage points—initially tethered Android devices, and ultimately a lightweight measurement app—to issue traceroutes toward stable external destinations (e.g., Google, Akamai, Cloudflare). These probes, annotated with GPS metadata, ensure traffic is treated as legitimate subscriber flows rather than external scans blocked at network borders. From these traces we directly identify verified Packet Gateways (P-GWs / UPFs) as the final in-AS hop before egress. An open design question remains: how many probes, and with what geographic distribution, are needed for representative coverage? This is a key area where we seek community feedback.

Phase 2: Feature extraction from ground truth

Verified P-GW IPs will be analyzed for consistent features, including reverse DNS naming conventions, IPv6 structural hints, BGP prefix characteristics, and latency signatures. These features form a descriptive model of what makes an IP address a likely mobile egress gateway.

Phase 3: Scalable inference (outside-in)

We apply heuristics learned from inside-out detection of P-GWs—such as naming patterns, IPv6 structure, BGP context, and RTT signatures—to large-scale public datasets. CAIDA's ITDK derives traceroutes to randomly selected destinations in each routed /24 prefix via Ark monitors [2], while M-Lab's traceroute data is generated using scamper as a sidecar tool during NDT measurements, tracing from the server to the client that initiated the test [11]. We acknowledge that many of these outside-in traceroutes are truncated or blocked at provider boundaries. Nevertheless, by matching partial signals from these traces against our ground-truth model, we can infer candidate egress routers likely to serve as P-GWs. This hybrid approach combines the broad coverage of outside-in data with the precision of inside-out validation.

Phase 4: Evaluation and longitudinal tracking

We evaluate the accuracy of scalable inference against our controlled ground-truth probes, measuring precision/recall for P-GW identification and city-level error for geolocation. To strengthen validation, we will also leverage the historical Mobiperf [13] dataset (2012–2020), which contains around 370,000 inside-out traceroutes, 40% of which originate from mobile networks (primarily U.S. providers). Although dated, this dataset provides valuable validation material and a longitudinal baseline for assessing how P-GW deployments evolved during the transition from 3G to LTE to early 5G. Once validated, we will repeat the pipeline across historical snapshots of ITDK and BGP data to track changes in P-GW deployment over time. To strengthen validation, we also plan to explore partnerships with third-party platforms (e.g., Cloudflare) that observe mobile egress traffic at scale.

4 CONCLUSION

Although we do not yet have results to present, *NetBreakout* lays the groundwork for a scalable, reproducible framework to illuminate mobile network egress infrastructure. By sharing our design early, we hope to gather community feedback that will sharpen our

methodology, strengthen validation, and guide the policy-relevant questions this work can address.

REFERENCES

[1] Bangladesh Telecommunication Regulatory Commission. 2024. Internet Subscribers in Bangladesh. https://btrc.gov.bd/site/page/347df7fe-409f-451e-a415-65b109a207f5/ Accessed: 2024-12-12.

[2] CAIDA, The Center for Applied Internet Data Analysis. 2022. Macroscopic Internet Topology Data Kit (ITDK). https://www.caida.org/catalog/datasets/internet-topology-data-kit. Accessed: August 3, 2025.

[3] Bradley Huffaker, Marina Fomenkov, and KC Claffy. 2014. DRoP: DNS-based router positioning. *ACM SIGCOMM Computer Communication Review* 44, 3 (2014), 5–13.

[4] International Telecommunication Union. 2023. Facts and Figures 2023: Internet Traffic. https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-traffic/ Accessed: 2024-12-12.

[5] Matthew Luckie, Amogh Dhamdhere, Bradley Huffaker, David Clark, and KC Claffy. 2016. Bdrmap: Inference of borders between IP networks. In *Proceedings of the 2016 internet measurement conference*. 381–396.

[6] Matthew Luckie, Bradley Huffaker, Alexander Marder, Zachary Bischof, Marianne Fletcher, and KC Claffy. 2021. Learning to extract geographic information from internet router hostnames. In *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*. 440–453.

[7] Matthew Luckie, Alexander Marder, Marianne Fletcher, Bradley Huffaker, and K Claffy. 2020. Learning to extract and use ASNs in hostnames. In *Proceedings of the ACM Internet Measurement Conference*. 386–392.

[8] Alexander Marder, Matthew Luckie, Amogh Dhamdhere, Bradley Huffaker, KC Claffy, and Jonathan M Smith. 2018. Pushing the boundaries with bdrmapit: Mapping router ownership at internet scale. In *Proceedings of the Internet Measurement Conference 2018*. 56–69.

[9] Alexander Marder and Jonathan M Smith. 2016. MAP-IT: Multipass accurate passive inferences from traceroute. In *Proceedings of the 2016 Internet Measurement Conference*. 397–411.

[10] Alexander Marder, Zesen Zhang, Ricky Mok, Ramakrishna Padmanabhan, Bradley Huffaker, Matthew Luckie, Alberto Dainotti, Alex C Snoeren, Aaron Schulman, et al. 2023. Access denied: Assessing physical risks to internet access networks. In *32nd USENIX Security Symposium (USENIX Security 23)*. 6877–6892.

[11] Measurement Lab (M-Lab). 2025. M-Lab Scamper Traceroute Dataset. https://measurementlab.net/tests/traceroute/. Accessed: August 3, 2025.

[12] United Nations. 2020. Advancing Human Wellbeing: A High-Level Political Forum on Sustainable Development. https://sdgs.un.org/sites/default/files/documents/26482HLPF_Advancing_human_wellbeing_BN_FINAL_1July2020.pdf Accessed: 2024-12-12.

[13] Ashkan Nikravesh, David R Choffnes, Ethan Katz-Bassett, Z Morley Mao, and Matt Welsh. 2014. Mobile network performance from user devices: A longitudinal, multidimensional analysis. In *International Conference on Passive and Active Network Measurement*. Springer, 12–22.

[14] Ohio State University College of Medicine. 2021. Internet Access as a Public Health Issue. (2021). https://medicine.osu.edu/news/internet-access-as-a-public-health-issue Accessed: 2024-12-12.

[15] Philipp Richter, Florian Wohlfart, Narseo Vallina-Rodriguez, Mark Allman, Randy Bush, Anja Feldmann, Christian Kreibich, Nicholas Weaver, and Vern Paxson. 2016. A multi-perspective analysis of carrier-grade NAT deployment. In *Proceedings of the 2016 Internet Measurement Conference*. 215–229.

[16] Rick Rojas, John McGee, and Others. 2020. When Nashville Bombing Hit a Telecom Hub, the Ripples Reached Far Beyond. *The New York Times* (Dec. 29 2020). https://www.nytimes.com/2020/12/29/us/nashville-bombing-telecommunications.html Accessed 2025-08-25.

[17] Quirin Scheitle, Oliver Gasser, Patrick Sattler, and Georg Carle. 2017. HLOC: Hints-based geolocation leveraging multiple measurement frameworks. In *2017 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 1–9.

[18] Neil Spring, Ratul Mahajan, and David Wetherall. 2002. Measuring ISP topologies with Rocketfuel. *ACM SIGCOMM Computer Communication Review* 32, 4 (2002), 133–145.

[19] Dhaka Tribune. 2023. Internet users in Bangladesh reach 131m as of 2023. https://www.dhakatribune.com/bangladesh/339218/internet-users-in-bangladesh-reach-131m-as-of-2023?utm_source=chatgpt.com Accessed: 2024-12-12.

[20] Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, Kimberly C Claffy, and Aaron Schulman. 2021. Inferring regional access network topologies: Methods and applications. In *Proceedings of the 21st ACM Internet Measurement Conference*. 720–738.