An Improved BGP Internet Graph for Optimizing Refraction Proxy Placement

Harshith Umesh
Red Hat
Boston, Massachusetts, USA
haumesh@redhat.com

Alden W. Jackson
Northeastern University
Boston, Massachusetts, USA
a.jackson@northeastern.edu

Abstract

Refraction Networking (RN) circumvents state-level censorship by embedding covert proxies within transit Autonomous Systems (ASes) that naturally lie on client-to-server paths. Selecting which ASes to recruit is a topology-aware optimization problem. However, existing AS-level maps often omit private peering relationships, CDN detours, and Internet Exchange Point (IXP) fabrics, leading to incomplete or inaccurate coverage estimates. In this work, we construct, what we believe is, the first censorship aware AS+IXP multigraph, combining traceroute derived forwarding paths (CAIDA), BGP-advertised AS adjacencies (RouteViews and RIPE RIS), and IXP membership data (CAIDA), all of which were collected in January 2025. Using this enriched graph, we conduct a detailed coverage analysis of DNS resolution paths for the five most censored countries (as identified by Censored Planet). We compute per-AS statistics, including usage frequency, unique path contribution, and cumulative coverage, to identify a minimal set of potential ASes for proxy placement. Our graph statistics align with known topologies from ASRank and CAIDA, validating the realism of our model. This work presents a scalable framework for censorship-aware Internet topology analysis, providing empirical insights for designing more resilient and targeted systems to circumvent censorship.

Keywords

Autonomous Systems, Internet Topology, Proxy Placement, Censorship, Refraction Networking, IXP

1 Introduction

Refraction networking proxies help circumvent sophisticated censorship by embedding traffic inside conventional Internet flows. Placing RN proxies effectively requires a thorough understanding of the locations of key Autonomous Systems, IXPs, and cross-AS connections. This paper demonstrates an enhanced model of the Internet topology, integrating BGP adjacency, IXP interconnections, and traceroute data to identify the minimal set of vantage ASes that maximize coverage for potential censored traffic.

Internet traffic routinely crosses dozens of independently managed networks (ASes) before reaching its destination. When a censor orders local Internet Service Providers (ISPs) to block access to a website, that order can be enforced at the AS level by poisoning DNS answers, injecting forged TCP resets, or silently dropping packets. These interventions create "chokepoints" that anti censorship systems must bypass. Refraction Networking is one of the most promising bypass strategies: instead of using a proxy with a well-known domain name, RN places a covert proxy inside a cooperative transit AS that lies on the forward path from the censored client to a benign "decoy" host. All traffic to the decoy passes through the covert proxy, making it extremely difficult for a censor to distinguish between censored and benign flows.

Selecting which ASes should host RN proxies is, therefore, in part a topology-driven optimization problem. The chosen deployment set should intercept the largest possible fraction of client-to-destination paths while minimizing costs and political risk. Unfortunately, the AS-level maps used in prior placement studies are incomplete. BGP tables omit (i) hidden local peering links inside Internet Exchange Points, (ii) content delivery detours where large CDNs tunnel traffic through leased prefixes, and (iii) IXP fabric details where hundreds of ASes share a Layer-2 switch but advertise only a single BGP next hop. Failing to consider these elements results in inaccurate coverage estimates and potentially suboptimal proxy placement.

This paper presents, to the best of our knowledge, the first censorship-aware AS+IXP multigraph that explicitly represents all three edge types relevant to RN deployment: public BGP adjacencies, private IXP connections, and traceroute-only links. We built the graph by fusing three large-scale measurement sources collected in January 2025: 13.6 M traceroutes from 14 globally distributed vantage points [3], 256 M BGP updates and RIBs from RouteViews [21] and RIPE RIS [19] collectors, and comprehensive IXP membership records from CAIDA [6] (collected from PeeringDB, Packet Clearing House (PCH), and Hurricane Electric (HE)).

The resulting graph contains 87,157 AS vertices, 1,588 IXP vertices, and 510,810 edges an order of magnitude richer than BGP only baselines. We couple this topology with five country-level DNS censorship datasets from Censored Planet [9] (Turkmenistan, China, Iran, Oman, Afghanistan) comprising 1.3 M domain queries to identify resolver ASes within censored regions and uncensored destination ASes elsewhere.

Our solution allows practitioners to test RN placement scenarios for any AS, set of ASes, or country mix. By closing long standing visibility gaps in Internet topology, our work enables the more effective and defensible deployment of refraction proxies, providing a reusable foundation for studies on routing security, fault tolerance, and traffic localization.

2 Background and Motivation

The Internet's logical "wiring diagram" is formed by ASes that exchange reachability information with the Border Gateway Protocol

Free and Open Communications on the Internet 2025(2), 51–59

(BGP). A precise picture of how those ASes interconnect is vital for tasks such as outage analysis, traffic engineering, and, more recently, anti-censorship systems. Yet three well-known obstacles keep the real AS-level topology partially hidden:

- Hidden (or "private-to-private") peering: Many AS pairs exchange traffic over unadvertised links inside IXPs or private facilities. These edges do not appear in BGP tables and therefore evade conventional topology crawlers.
- Content Delivery Networks (CDNs): Large CDNs frequently lease address space and tunnel traffic in ways that make origin AS identification ambiguous. CDN localisation also changes the physical path used by a flow, complicating inferences derived solely from BGP.
- IXP complexity: Modern IXPs can contain tens to hundreds of member ASes connected via switching fabrics or route-servers. A BGP relationship between two such members does not always imply they share a direct physical link; conversely, many Layer-2 adjacencies are invisible to BGP.

These blind spots are particularly damaging to RN, a censorship evasion technique that hides proxies inside cooperative transit ASes. The effectiveness of an RN deployment hinges on refracting a significant fraction of user-to-uncensored destination paths. To date, site selection papers (e.g. TapDance, Telex [23] [22]) assume a clean BGP-only graph and therefore risk under- or overestimating achievable path coverage.

Our motivation is to find AS and/or IXP vantage points that see the largest fraction of Source–Destination (censored user → uncensored domain) paths for placing Refraction Proxies. By computing shortest AS-level paths between ASes inside censored regions to uncensored destination ASes, we can rank transit ASes by (i) raw usage, (ii) coverage gains, and (iii) unique paths traversed. Preliminary results show that, for example, for Iran, instrumenting only 4–10 ASes yields 50–75 percent path interception, which is not possible with earlier models.

3 Related Work

Our work intersects three principal areas of research: Internet topology mapping, censorship measurement and analysis, and the strategic placement of proxies or monitors within the network.

Substantial work has been dedicated to uncovering the structural layout of the Internet. Motamedi et al. [13] provide a comprehensive taxonomy of topology discovery techniques across multiple resolutions, interface, router, Point of Presence (PoP), and AS-level, highlighting the challenges of incomplete visibility due to the Internet's decentralized nature . Nur and Tozal [24] extend traditional AS-level graphs to multigraphs by incorporating multiple inter-AS connections, which significantly improve the fidelity of Internet topology models, especially for cross border analysis. Nur [14] further explores this in the context of AS-level graphs and multigraphs, demonstrating the role of parallel edges in characterizing realistic topologies.

Ahmad and Guha [1] investigate the influence of IXPs on topology evolution, showing how IXPs reshape AS-level connectivity and affect peering strategies . While they highlight the impact of IXPs on latency and topology evolution, our study leverages IXPs as connectivity hubs, we find that strategically located IXPs can serve as "shortcuts" through which a single proxy instance may intercept paths from multiple ASes. This leverages the structural redundancy of IXPs to create a deployment advantage.

We extend the Cross-AS multigraph concept proposed by Nur and Tozal [14] by constructing an AS+IXP multigraph that incorporates explicit IXP membership into the topology. This allows us to evaluate IXPs as active relay points for refraction networking proxies, an aspect Ahmad and Guha [1] acknowledged, but did not model structurally.

Jackson et al. [11] address the optimal placement of passive monitors within the AS topology, treating it as a coverage maximization problem. Their results, based on Skitter [2] and RouteViews [21] data, show that deploying shallow monitors across many ASes offers better coverage than deeply instrumenting a few Tier-1s. Our greedy coverage by summation algorithm builds upon this philosophy. We rank ASes by their marginal contribution to DNS resolution paths, and demonstrate that in most countries, deploying proxies in the top 1–12 ASes suffices to intercept 75% of all uncensored DNS paths, mirroring the breadth-first logic of Jackson et al., but with a country-specific and censorship-aware lens.

The Censored Planet platform [20] has become a foundational infrastructure for Internet-wide censorship monitoring. It continuously collects DNS and HTTP interference data, as shown by Raman et al. [17], and emphasizes robust AS and geolocation attribution to avoid classification errors. We adopt their resolver/answer distinction methodology, filter only non-censored resolution paths, and integrate this data directly into our graph based path analysis, ensuring that only viable, censorship-free endpoints inform our proxy placement evaluations.

Several notable anti-censorship systems such as Telex [23], TapDance [22], and Rebound [10] relocate the proxy logic into the core of the Internet. These systems aim to bypass IP and SNI-based filtering by embedding proxies inside ISP infrastructure. However, a critical and largely unanswered question remains: Where should such cooperating routers be placed to achieve the maximum effect? Our work directly addresses this gap. By computing path coverage frontiers across five highly censored countries, we offer the first large scale, empirical analysis of AS-level placements suited for refraction networking, which can directly inform the deployment of future RN proxies.

Whereas previous work has either (i) mapped Internet topology, (ii) measured censorship behavior, or (iii) designed probe resistant protocols, our contribution is integrative: we build a structurally validated AS+IXP multigraph, annotate it with real censorship-aware resolution paths, and derive deployment sets of ASes and IXPs that meet coverage thresholds. This closes a key operational gap between abstract network measurement and practical circumvention deployment planning, offering a foundation for the next generation of refraction networking systems.

4 Methodology

Our methodology transforms large-scale, heterogeneous Internet measurements into ranked deployment targets for refraction networking. Figure 5 outlines the end-to-end pipeline.

Our approach combines control-plane (BGP) and data-plane (traceroute) datasets with censorship measurement datasets to

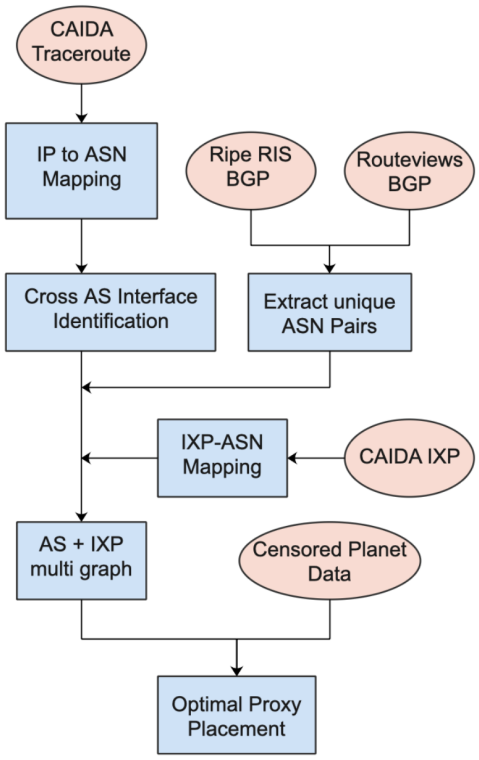

Figure 1: Workflow for Optimal Proxy Placement

model routing behavior and identify Autonomous Systems (ASes) that are strategically positioned to serve as effective points for circumventing censorship. The IXP dataset serves as auxiliary topology metadata, enriching both control and data plane interpretations by revealing where AS–AS connections may occur indirectly via shared IXP memberships.

All datasets were collected around January 22, 2025, ensuring temporal consistency across measurement layers. The DNS censorship data, obtained from the Censored Planet, was collected on January 15, 2025, the closest available dataset to the others.

The methodology proceeds through six algorithmic steps. The steps are represented as blue boxes in Figure 5. We begin by resolving every traceroute hop to its origin AS using a longest-prefix trie built from CAIDA's prefix-to-AS dataset (Algorithm 1). Next, we discover cross-AS router interfaces directly visible in traceroute paths (Algorithm 2) and augment these edges with adjacent AS pairs extracted from 256 million BGP updates drawn from RIPE RIS and RouteViews (Algorithm 3). We then model IXP memberships by mapping IXPs to their member ASes (Algorithm 4). These four data streams are fused into an AS+IXP multigraph (Algorithm 5). Finally, for each country specific censorship dataset from Censored Planet, we replay resolver→destination flows over this graph and apply a greedy set cover heuristic to identify the minimum set of transit ASes required to intercept 25%, 50%, and 75% of all uncensored paths (Algorithm 6). The resulting ranked lists constitute our recommended refraction proxy placements. The following subsections elaborate on each of these steps.

4.1 Step 1: IP to ASN Mapping

We begin by converting IP-level traceroute data [3] from the CAIDA prefix probing dataset into AS-level paths using a prefix trie constructed from CAIDA's prefix-to-AS mapping dataset [7]. We apply the longest prefix match to each IP address in the traceroute hops. We use a prefix trie because it gives fast, memory-efficient longest-prefix matching with $O(L)$ lookup time, where L is the prefix length in bits. The dataset comprises 13,562,991 traceroute paths, encompassing 944,224 unique IP addresses. In total, 149,873,599 IP hops were processed, out of those, 259,933 IP addresses (0.173%) with no corresponding ASNs were marked "NA".

This mapping is performed on traceroutes collected from 14 global vantage points, ensuring coverage across regions. We use the `scamper` [8] tool to efficiently parses CAIDA traceroute `warts` files into hop-level IP paths for subsequent processing. The resulting AS annotated traceroutes provide the foundation for inter-AS analysis.

Algorithm 1: IP-to-ASN Mapping

Input: `pfx2as_file` – Prefix-to-AS mappings
`traceroute_file` – Traceroute paths
Output: IP-to-ASN mapping

```
trie ← LOAD_PFX2AS_INTO_TRIE(pfx2as_file);
ip_set ← PARSE_TRACEROUTE(traceroute_file);
ip_asn_map ← ∅;
foreach ip ∈ ip_set do
    asn ← trie.search(ip);
    if asn is None then
        asn ← "NA";
    ip_asn_map[ip] ← asn;
```

4.2 Step 2: Cross-AS Interface Identification

From the annotated traceroute paths, we identify inter AS transitions by scanning adjacent hops. If two consecutive IPs belong to different ASNs, the hop is considered a cross-AS interface. This step yielded 1,086,764 unique cross-AS pairs, capturing AS level forwarding behaviors observed in the data plane. These interfaces are later correlated with IXP data and BGP peering relationships to enrich connectivity analysis.

4.3 Step 3: Extract unique ASN Pairs

To complement data-plane interfaces, we extract AS adjacencies from BGP datasets. Specifically, we use BGP Multi Route Table (MRT) dump files collected from Routeviews [21] and RIPE RIS [19] multi-hop collectors. MRT records encapsulate BGP `UPDATE` or `TABLE DUMP` messages. Multi-hop route collectors collect data from peers via BGP multi-hop sessions. Multi-hop sessions let a single collector ingest feeds from multiple routers worldwide without being physically present at every site. For this study we parsed BGP MRT files collected from the multi-hop collectors `rv2, rv3, rv4, rv5`, and `rv7` operated by RouteViews and `rrc00, rrc24`, and `rrc25` along with the single-hop collector `rrc13` operated by RIPE RIS. The combined BGP dataset contains: 132,675,689 RIPE BGP records and 123,802,034 RouteViews BGP records. We use the

Algorithm 2: Identify Cross-AS Interfaces

Input: `traceroute_file` – CAIDA traceroute file, where each trace is an ordered list of IP addresses
`ip_asn_map` – IP addresses to ASN Mapping
Output: Set of ASN pairs representing cross-AS neighbors

```
cross_asns ← ∅;
foreach trace ∈ traceroute_file do
    for i ← 0 to |trace| − 2 do
        ip1 ← trace[i];
        ip2 ← trace[i+1];
        asn1 ← ip_asn_map.get(ip1, "NA");
        asn2 ← ip_asn_map.get(ip2, "NA");
        if asn1 ≠ "NA" and asn2 ≠ "NA" and asn1 ≠ asn2
          then
            cross_asns ← cross_asns ∪ {(asn1, asn2)};
return cross_asns;
```

bgpdump tool [18] to decode and extract the routing tables from the MRT files.

We extract unique adjacent AS pairs from the AS path field. Each pair implies at least one peering or transit link between the ASes. This step provides the control-plane perspective of AS connectivity and is merged with traceroute-based interfaces in Step 5.

Algorithm 3: EXTRACT_UNIQUE_AS_PAIRS

Input: `BGP_MRT_file` – RIPE/Routeviews MRT files
Output: Sorted set of unique adjacent AS pairs

```
as_pairs ← ∅;
foreach line in BGP_MRT_file do
    as_path_str ← Extract AS Path;
    as_numbers ← as_path_str.split();
    for i ← 0 to length(as_numbers) −2 do
        as1 ← int(as_numbers[i]);
        as2 ← int(as_numbers[i+1]);
        if as1 ≠ as2 then
            as_pairs.insert((as1, as2));
```

4.4 Step 4: IXP-ASN Mapping

We leverage CAIDA's IXP dataset [6] to map ASNs to IXPs. This step produces two bidirectional mappings:
`ixp_to_asns`: maps each IXP ID to its member ASNs
`asn_to_ixps`: maps each ASN to the IXPs it connects to.

Out of 1,588 total IXPs, 1,464 have at least one associated ASN, while 124 IXPs remain isolated. This information is critical for modeling IXP-mediated AS interconnectivity and is incorporated into the multigraph.

Algorithm 4: Build IXP–ASN Mappings

Input: `ix_asns_file` – CAIDA IXP file with fields `ix_id` and `asn`
Output: IXP→ASNs mapping and ASN→IXPs mapping

```
ixp_to_asn ← empty map (ix_id → set of ASNs);
asn_to_ixp ← empty map (asn → set of ix_ids);
foreach line in ix_asns_file do
    Extract ix_id, asn;
    ixp_to_asn[ix_id] ← ixp_to_asn[ix_id] ∪ {asn};
    asn_to_ixp[asn] ← asn_to_ixp[asn] ∪ {ix_id};
```

4.5 Step 5: AS+IXP Multigraph

To identify the most effective sites for deploying proxies, we require a topology representation that includes all layers at which ASes exchange traffic: BGP connections, hidden peering links visible only in traceroute, and IXP interconnections. The following algorithm turns heterogeneous datasets into a single multigraph g, enriched with semantic labels that permit filtering "BGP edges", "IXP edges", or "traceroute-only" edge types at will. The resulting graph is comprised of vertices which represent ASNs or IXPs and edges which represent:
1. Direct BGP peering relationships (from Step 3)
2. Cross-AS traceroute observations (from Step 2)
3. AS–IXP memberships (from Step 4)

All graph operations are performed with the `graph-tool` Python library [16]. We selected `graph-tool` because it supports $O(1)$ adjacency look-ups and duplicate-edge checks, and offers a highly optimised C++ function for the shortest-path algorithm used in Step 6.

The final graph contains 88,621 vertices and 510,810 edges, including 87,157 AS nodes and 1,588 IXP nodes. The graph is annotated with vertex properties (`node_type, asn, ixp_id`) and edge properties (`edge_type, ixp_list`) and forms the structural basis for the shortest-path and coverage analysis in Step 6.

4.6 Step 6: Optimal Proxy Placement

This step translates the AS-level multigraph into actionable guidance on where to deploy refraction proxies. To assess censorship-aware routing and proxy placement, we use the Censored Planet DNS censorship dataset for the five countries most censored, ranked by Censored Planet [9].

For each target country, Censored Planet performs DNS resolution probes across a large set of domains. It reports whether the resolution was successful or blocked, as well as the ASN of the resolver and the ASN associated with the resolved domain. We focus on the top five countries with the highest observed levels of DNS-based censorship, as ranked by Censored Planet: Iran, China, Turkmenistan, Oman, and Afghanistan. For each country, we extract three key metrics: the number of domains probed, the percentage of domains that were censored, and the sets of ASNs observed as resolvers and uncensored domain providers, shown in Table 1.

For each country, we extract the Resolver ASNs and Uncensored ASNs. Resolver ASes are ASes that answer DNS look-ups from within a country, and Uncensored ASes are ASes hosting domains that return non-censored replies.

Any domain classified as censored is discarded, ensuring that we only consider routes that are not blocked by the country. We compute the shortest path between a resolver and uncensored ASNs, using only the shortest AS path BGP policy, as individual AS BGP policies are not visible from AS path data. Along each path, we record the intermediate ASes and IXPs. We exclude any path where a resolver AS appears in the computed shortest path. We then calculate:

- Usage count: Frequency of each AS/IXP across all paths
- Coverage percentage: Percentage of all paths that traverse a given AS
- Unique contribution: Number of client to destination paths that only one AS can intercept once the ASes already instrumented with a proxy are taken into account. It is derived by performing a greedy sweep where paths are first assigned to the AS with the highest Usage count, and any path that remains uncovered is then checked against the next AS, and so on. Unique coverage allows us to prioritize ASes that add new, non-overlapping visibility into uncensored paths, rather than duplicating coverage already obtained.
- Cumulative coverage: We calculate the cumulative coverage achieved by the top-k ASes using a greedy coverage approach by summation. The ASes are sorted by usage counts in decreasing order and accumulate paths until the running total exceeds 25%, 50% or 75% of all paths.

These metrics guide proxy placement, where ASes with high path coverage and unique contribution are prioritized for deploying censorship circumvention infrastructure.

Algorithm 5: Build AS–IXP Graph

Input: `asn_to_ixp` – ASN to set of IXP IDs map
`ixp_to_asn` – IXP ID to set of ASNs map
`asn_neighbors_traceroute` – traceroute_cross_asns
`asn_neighbors_bgp_rv` – routeviews as_pairs
`asn_neighbors_bgp_ripe` – ripe ris as_pairs
Output: AS + IXP Graph g

asn_neighbors_bgp ← merge(asn_neighbors_bgp_rv, asn_neighbors_bgp_ripe);
CREATE graph g with:
- vertex_properties: `node_type`, `asn`, `ixp_id`
- edge_properties: `edge_type`, `ixp_list`

foreach *ASN in asn_neighbors_traceroute or asn_neighbors_bgp* **do**
  ADD vertex to g with node_type = "AS", asn = ASN;
foreach *IXP ID in ixp_to_asn* **do**
  ADD vertex to g with node_type = "IXP", ixp_id = IXP ID;
foreach *(asn1, asn2)* ∈ *asn_neighbors_bgp* **do**
  ADD edge between asn1 and asn2 with edge_type = "BGP";
foreach *(asn1, asn2)* ∈ *asn_neighbors_traceroute* **do**
  shared_ixps ← ixp_shared(asn1, asn2, asn_to_ixp);
  if *shared_ixps* ≠ ∅ **then**
    foreach *ixp* ∈ *shared_ixps* **do**
      ADD edge from asn1 to ixp and asn2 to ixp with edge_type = "IXP";
    ADD edge between asn1 and asn2 with edge_type = "TRACEROUTE_IXP";
  else
    ADD edge between asn1 and asn2 with edge_type = "TRACEROUTE_ONLY";
return as_ixp_graph.gt;

Algorithm 6: Coverage Analysis & Optimal proxy placement for a censoring country

Input:
AS+IXP topology graph g: `as_ixp_graph.gt`
Per-country DNS censorship data: `satellite-dns.json` from Censored Planet
Output: Coverage and usage metrics of ASes, IXPs participating in DNS resolution paths

Parse `country/satellite-dns.json` to extract:;
  `resolvers` ← set of resolver ASNs;
  `uncensored` ← set of uncensored ASNs from DNS answers;
Initialize `asn_vertex_map` mapping each ASN to its vertex in *g*;
Initialize `ixp_vertex_map` mapping each IXP ID to its vertex in *g*;
Initialize `path_map` ← ∅, `total_paths` ← 0;
foreach *resolver ASN r* **do**
  foreach *uncensored ASN u* **do**
    Get vertices v_r, v_u from `asn_vertex_map`;
    Compute shortest path *P* from v_r to v_u in *g*;
    if *valid path P found* **then**
      `total_paths++`;
      Extract intermediate AS/IXP nodes and store in `path_map[path_id]`;
      Increment `path_id`;
Initialize `usage_counts` ← ∅;
foreach *path in* `path_map` **do**
  foreach *ASN in path* **do**
    `usage_counts[ASN]++`;
  foreach *IXP in path* **do**
    `usage_counts[IXP]++`;

4.7 Limitations

Note that it is possible for transit ASes of a censoring country, those that do not host resolvers themselves, to show up in our shortest

path computations. This is because the Censored Planet dataset records only resolver ASes. We identify and flag these cases in a post-processing step, as shown in our results.

One limitation of our method is the derivation of the AS-level shortest path. We construct the shortest path from our AS+IXP multigraph with a preference for traceroute edges. If such a path is absent, we fall back to the constructed BGP shortest path. This procedure implicitly assumes that ASes always forward traffic along the shortest hop count, disregarding real-world BGP routing policies. As a consequence, some of the computed paths may be topologically valid yet operationally infeasible, which could lead to overestimating coverage for certain ASes and underrepresenting policy-preferred routes.

Our AS+IXP multigraph construction is subject to limitations inherent to all Internet measurement studies. Missing edges may occur when vantage points fail to see specific AS paths due to restricted BGP visibility, inadequate traceroute coverage, or filtering by resolvers or upstream providers. Conversely, false edges may be introduced through measurement artifacts such as IP aliasing errors in traceroute, misinferred AS relationships, or stale or misconfigured BGP announcements.

The impact of missing an edge typically results in an underestimation of path reachability, causing certain ASes that may experience considerable traffic to be undervalued. This may lead us to miss particular high-impact perspectives in our RN placement suggestions. On the other hand, incorrectly adding an edge could exaggerate the perceived centrality or reach of an AS, potentially leading us to recommend proxy deployments in locations that do not offer proper coverage of censored paths.

To mitigate these effects, we rely on multiple complementary data sources collected on the same day (CAIDA traceroute, BGP from RIPE RIS and RouteViews, and IXP membership data from CAIDA), which allows us to cross-check visibility and minimize the chances of systemic error. However, we acknowledge that no dataset offers perfect ground truth, and small inaccuracies can propagate through path calculations.

5 Results

Our AS+IXP multigraph allows a detailed analysis of AS-level connectivity by categorizing inter-AS edges into four types: BGP, traceroute-only, IXP, and traceroute-IXP.

We use our graph to generate ranking lists for ASes based on:
1. Total number of AS neighbors
2. Number of BGP edges
3. Number of private/traceroute only edges
4. IXP connectivity (both direct and inferred).

5.1 Constructed Topology Map Statistics

In our ranking, AS3356 (Level 3 parent, LLC) emerges as the most connected AS with 14,209 total neighbors, followed by AS174 (Cogent Communications) and AS6939 (Hurricane Electric). ASes AS3356 and AS174 also top the list for traceroute-only edges (7,344 and 7,148, respectively), highlighting their significant presence in unadvertised or peering relationships not visible via BGP. However, they are not the top ASes by BGP neighbors, that spot is taken by AS6939.

Moreover, AS6762 (Telecom Italia Sparkle) ranks highest in traceroute-IXP edges (1,406 connections), indicating extensive indirect inter-AS connectivity via shared IXPs that are captured only through traceroute observations.

In terms of IXP membership, AS13335 (Cloudflare) has the highest IXP degree with 292 direct IXP edges, emphasizing its hyper-connectivity across exchange points. ASes AS16509 (Amazon) and AS6939 follow.

To validate the structural fidelity of our constructed AS+IXP multigraph and assess how accurately it models real-world Internet topology, we compared key connectivity metrics derived from our graph with those from authoritative external datasets, such as the ASRank [4] and the CAIDA IXP dataset [6]. For example, AS3356 (Level 3 Parent, LLC), which ranks first in ASRank due to its large customer cone size, appears as the top AS in our graph by total number of AS neighbors. Similarly, AS6939 (Hurricane Electric), which has the highest transit degree globally, tops our ranking for the number of BGP edges. Furthermore, AS13335 (Cloudflare), with its extensive peering and CDN presence, has the highest number of IXP edges in our graph, in agreement with its listings in the CAIDA IXP membership dataset. These correlations provide strong evidence that our AS+IXP multigraph is a representative model of the global Internet topology.

5.2 Case Study: Iran

We evaluate DNS resolution paths for Iran using our AS+IXP multigraph and analyze ASes for their strategic potential in proxy placement. The analysis is based on 22,799 successful resolver to uncensored AS paths that we generate (57 resolver ASNs X 400 uncensored ASNs).

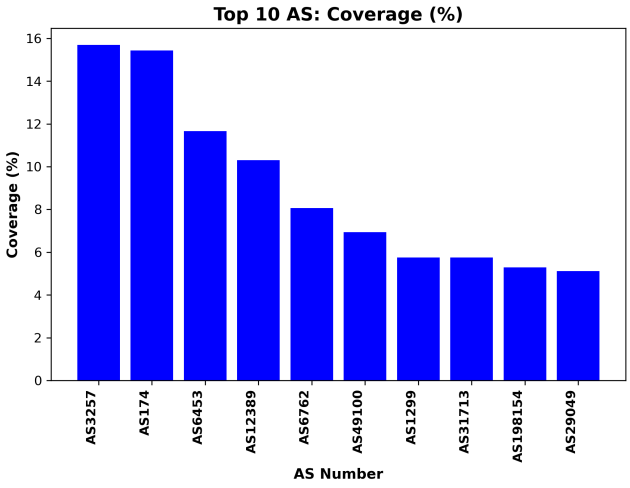

Figure 2: Top 10 AS: Coverage Percentage for Iran

We begin our analysis by identifying the ASes most frequently traversed across Iranian DNS resolution paths. Figure 2 shows that AS3257 (GTT Communications, US) and AS174 (Cogent Communications, US) appear in approximately 15.7% of all paths each, making them the most influential transit providers for Iranian DNS traffic. These are followed by AS6453 (TATA Communications, US),

AS12389 (PJSC Rostelecom, RU), and AS6762 (Telecom Italia Sparkle, IT), all of which are well-known Tier-1 or Tier-2 transit networks. The prominence of these ASes underscores their centrality in routing uncensored traffic from Iran to external destinations, making them prime candidates for strategic proxy placement.

To aid interpretation of the AS numbers shown along the x-axis in the figures, a mapping of ASNs to organization names, countries, and the number of AS adjencies is provided in the Appendix. Our tool uses the AS to organization dataset by CAIDA [5] to map ASNs to AS names and organizations.

To quantify the number of ASes required to account for an increasing share of all resolution paths, we perform a greedy cumulative coverage ranking. In this strategy, each path is attributed to the highest-ranked AS it traverses. Figure 3 shows that the top 5 ASes cumulatively cover over 59% of all paths, and the top 10 ASes account for 76.6%. This suggests that a relatively small number of well-connected ASes can capture a majority of the resolution paths.

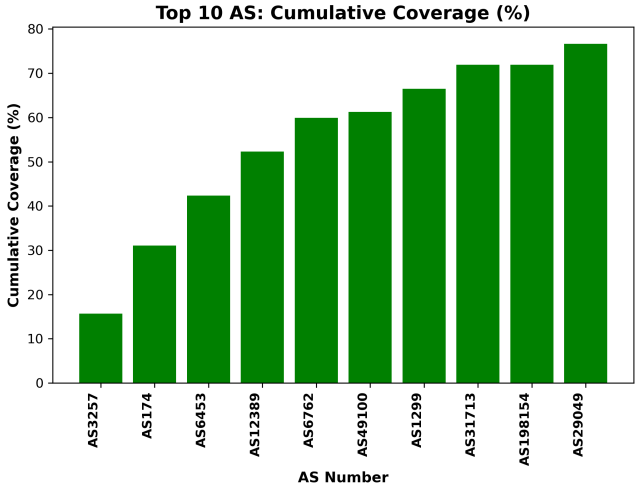

Figure 3: Top 10 AS: Cumulative coverage % for Iran

To highlight redundancy among high-ranked ASes and identify nodes that offer exclusive routing visibility, we calculate unique coverage. Figure 4 shows that AS3257 and AS174 retain nearly all their usage as unique paths, underscoring their exclusivity and importance. In contrast, AS12389 (PJSC Rostelecom, RU), while present in 2,350 paths, contributes only 1475 unique paths, meaning 37% of its routes are already captured by more central ASes. These results inform a minimal yet effective selection of ASes for proxy placement, favoring those that contribute the most non-overlapping path visibility.

Finally, we examine IXP involvement in DNS resolution paths. While ASes dominate path-level visibility, certain IXPs play supportive roles in inter-AS connectivity. Among 1,588 IXPs in our dataset, only a few are frequently traversed. DE-CIX Frankfurt (Germany) is the most visible, appearing in 80 paths, followed by NL-IX (Netherlands) with 3 paths, and DE-CIX Istanbul, AMS-IX, and NIXI (India) with one path each. The relatively low frequency of IXP-mediated paths suggests that most Iranian resolution paths either

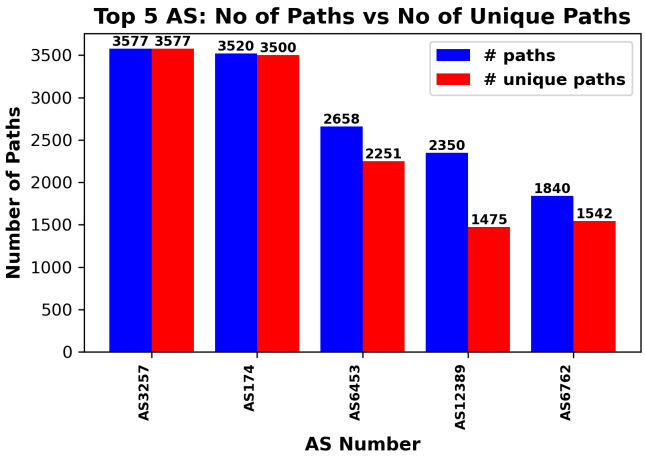

Figure 4: Number of Paths vs unique paths observed by Top 5 ASes for Iran

use direct AS-to-AS links or pass through private peerings, possibly due to regional infrastructure limitations or policy constraints.

5.3 Real World Considerations

While our study frames RN proxy placement as a topology-driven optimization problem, we recognize that real-world deployments involve additional constraints beyond structural coverage. Factors such as cost, legal jurisdiction, and the willingness of ASes to support refraction networks are crucial in practice. Our analysis provides a foundational view of which ASes could be leveraged for high coverage, but actual deployment may be limited to a smaller, cooperative set of candidates.

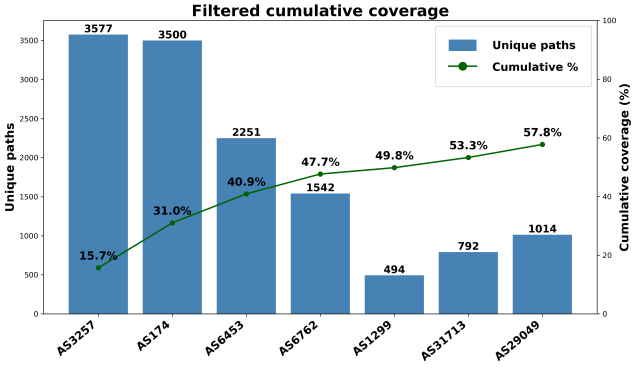

Figure 5: Filtered cumulative coverage for unique paths (after excluding uncooperative ASes)

Our framework can be flexibly adapted to these scenarios by allowing deployers to exclude uncooperative ASes from consideration. For example, in the case of Iran, if major ASes in countries such as Russia (AS12389) and Iran (AS49100, AS198154 - Note these are transit ASes and not resolver ASes), are deemed uncooperative, they can be removed from the deployment pool. The cumulative coverage for unique paths can then be recomputed to assess how

much visibility can still be retained using only cooperative ASes. As shown in Figure 5, even after removing three such ASes, several viable candidates remain, offering a practical starting point for deployment planning under geopolitical constraints.

Figures 3 and 5 depict our greedy deployment strategy, where ASes are ordered by their unique-path counts, and the curve shows the cumulative coverage obtained by adding them one at a time. Because path sets overlap, this sequence is not always the mathematically optimal combination for each k(k=1..N). We choose greedy order because operators would likely roll out proxies incrementally and evaluate each new site's marginal benefit.

5.4 Multi-Country Comparison

We extend our analysis to five censored countries — Iran, China, Turkmenistan, Oman, and Afghanistan, to assess how network centralization affects proxy placement strategies. For each country, we determine the number of ASes required to achieve 25%, 50%, and 75% coverage of all DNS paths. Figure 6 shows that China requires 12 ASes to achieve 75% coverage, indicating a highly fragmented routing topology. In contrast, Turkmenistan requires only one AS, reflecting an extreme level of centralization. Afghanistan requires 5 ASes, placing it between these two extremes. These differences reveal that censorship circumvention strategies must be tailored to each country's network structure, with more centralized regimes offering easier but potentially riskier opportunities for proxy placement.

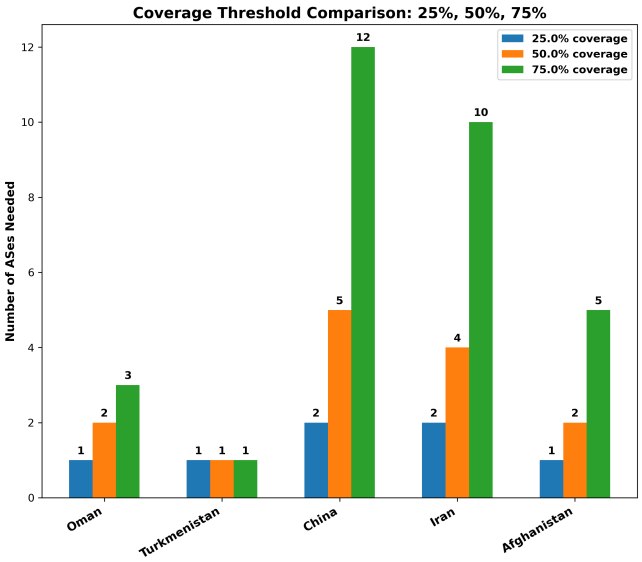

Figure 6: Multi-country Comparison

5.5 Stability of our Results

Prior longitudinal studies indicate that the tier I/II ASes similar to the ones our framework selects, change very little over months or even years. Oliveira et al. [15] show that almost all churn occurs at the customer edge, whereas large transit providers adjust their peering only gradually. The authors explicitly note that the core's vertex set of ASes and its high-capacity links remain "remarkably persistent" across consecutive quarterly snapshots. Liu et al. [12] broaden the study period to 1998–2013 and, after decomposing each yearly graph into its structural components, report that 95% of the core-to-core links remain unchanged. Together, these findings imply that an AS that intercepts a large share of paths today is likely to hold a similar position several months from now, lending strong support to the temporal robustness of our RN proxy placement recommendations.

6 Conclusion

In this paper, we present a comprehensive methodology for constructing an AS+IXP multigraph and applying it to censorship-aware routing analysis. By integrating control-plane data (BGP), data-plane observations (traceroutes), auxiliary IXP memberships dataset, and DNS censorship measurements, we create a unified multigraph that models real-world inter-domain connectivity with high fidelity. We validate the realism of this graph using edge type statistics and cross-reference with ASRank [4] and CAIDA [6] to confirm consistency with known Internet topological properties.

Through an extensive analysis of DNS resolution paths in heavily censored countries, we identify a small set of ASes that contribute disproportionately to routing visibility. Using metrics such as usage frequency, unique path contribution, and cumulative coverage, we propose a data-driven approach to optimal proxy placement. Our findings reveal structural differences among countries, for example, centralized routing in Turkmenistan versus more distributed fabric in China and Iran, which have direct implications for designing resilient circumvention strategies.

In general, our framework offers a scalable and extensible model for analyzing inter-domain routing under censorship, and our results contribute both empirical insights and methodological tools for future research in network resilience, policy enforcement, and freedom of information technologies.

7 Future Work

While our work identifies ASes and IXPs most suitable for deploying RN proxies, future work can extend this framework to optimize proxy placement for global coverage rather than just individual countries, minimizing the number of RN proxies needed while maximizing the number of censoring countries covered.

Acknowledgments

This work was supported by the Master's Research Apprenticeship program at Northeastern University Khoury College of Computing Sciences. We are grateful to CAIDA for access to the most recent year of Ark data. We are also grateful to RIPE, RouteViews, ASRank, and Censored Planet projects for publicly sharing their datasets. The authors thank Ms Aneesha AV for their support throughout this research. We also thank Professors David Choffnes and Rajagopal Venkatesaramani, as well as all the anonymous reviewers, for their helpful feedback. We thank Red Hat for their support.

References

[1] Mohammad Zubair Ahmad and Ratan Guha. 2010. Impact of Internet exchange points on Internet topology evolution. In *IEEE Local Computer Network Conference*. 332–335. https://doi.org/10.1109/LCN.2010.5735736

[2] CAIDA. 2008. Skitter Topology Dataset. https://www.caida.org/catalog/datasets/skitter-aslinks-dataset/.

[3] CAIDA. 2025. Ark IPv4 Prefix-Probing Dataset. https://doi.org/10.21986/CAIDA.DATA.ARK-IPV4-PREFIX-PROBING. https://doi.org/10.21986/CAIDA.DATA.ARK-IPV4-PREFIX-PROBING Accessed: 2025-01-21. Dates used: January 2025.

[4] CAIDA. 2025. AS Rank: A ranking of Autonomous Systems (AS) in the Internet. https://doi.org/10.21986/CAIDA.DATA.AS-RANK.

[5] CAIDA. 2025. AS to Organizations Mappings Dataset. https://catalog.caida.org/dataset/as_organizations. https://doi.org/10.21986/CAIDA.DATA.AS-ORGANIZATIONS Accessed: 2025-01-21. Dates used: January 2025.

[6] CAIDA. 2025. Internet eXchange Points Dataset. https://doi.org/10.21986/CAIDA.DATA.IXPS. https://doi.org/10.21986/CAIDA.DATA.IXPS Accessed: 2025-01-21. Dates used: January 2025.

[7] CAIDA. 2025. Routeviews Prefix to AS mappings Dataset for IPv4 and IPv6. https://www.caida.org/catalog/datasets/routeviews-prefix2as/. Accessed: 2025-01-21. Dates used: January 2025.

[8] CAIDA. 2025. scamper: A Network Measurement Tool. https://doi.org/10.21986/CAIDA.SOFTWARE.SCAMPER. https://doi.org/10.21986/CAIDA.SOFTWARE.SCAMPER Accessed: 2025-01-21. Dates used: January 2025.

[9] Censored Planet. 2025. Censored Planet: An Internet-wide, Longitudinal Censorship Observatory. https://censoredplanet.org/. Accessed: 2025-01-21. Dates used: January 2025.

[10] Daniel Ellard, Christine Jones, Victoria Manfredi, W. Timothy Strayer, Bishal Thapa, Megan Van Welie, and Alden Jackson. 2015. Rebound: Decoy routing on asymmetric routes via error messages. In *2015 IEEE 40th Conference on Local Computer Networks (LCN)*. 91–99. https://doi.org/10.1109/LCN.2015.7366287

[11] Alden W. Jackson, Walter Milliken, Cesar A. Santivanez, Matthew Condell, and W. Timothy Strayer. 2007. A Topological Analysis of Monitor Placement. In *Sixth IEEE International Symposium on Network Computing and Applications (NCA 2007)*. 169–178. https://doi.org/10.1109/NCA.2007.3

[12] Xiao Liu, Jinfa Wang, Wei Jing, Menno de Jong, Jeroen S Tummers, and Hai Zhao. 2018. Evolution of the Internet AS-level topology: From nodes and edges to components*. *Chinese Physics B* 27, 12 (dec 2018), 120501. https://doi.org/10.1088/1674-1056/27/12/120501

[13] Reza Motamedi, Reza Rejaie, and Walter Willinger. 2015. A Survey of Techniques for Internet Topology Discovery. *IEEE Communications Surveys & Tutorials* 17, 2 (2015), 1044–1065. https://doi.org/10.1109/COMST.2014.2376520

[14] Abdullah Yasin Nur. 2021. Analysis of Autonomous System Level Internet Topology Graphs and Multigraphs. In *2021 International Symposium on Networks, Computers and Communications (ISNCC)*. 1–7. https://doi.org/10.1109/ISNCC52172.2021.9615677

[15] Ricardo Oliveira, Beichuan Zhang, and Lixia Zhang. 2007. Observing the evolution of Internet AS topology. *ACM SIGCOMM Computer Communication Review* 37, 313–324. https://doi.org/10.1145/1282380.1282416

[16] Tiago P. Peixoto. 2014. The graph-tool python library. *figshare* (2014). https://doi.org/10.6084/m9.figshare.1164194

[17] R Raman, A Virkud, S Laplante, V Fortuna, and R Ensafi. 2023. Advancing the Art of Censorship Data Analysis. *Pet Symposium* (2023). https://doi.org/paper/2023_raman_r_pet_symposium

[18] RIPE NCC. 2025. bgpdump: A BGP Table Dump Parser. https://github.com/RIPE-NCC/bgpdump. Accessed: 2025-01-21. Dates used: January 2025.

[19] RIPE NCC. 2025. RIPE Routing Information Service (RIS). https://ris.ripe.net/. Accessed: 2025-01-21. Dates used: January 2025.

[20] Ram Sundara Raman, Prerana Shenoy, Katharina Kohls, and Roya Ensafi. 2020. Censored Planet: An Internet-wide, Longitudinal Censorship Observatory. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security* (Virtual Event, USA) *(CCS '20)*. Association for Computing Machinery, New York, NY, USA, 49–66. https://doi.org/10.1145/3372297.3417883

[21] University of Oregon Route Views Project. 2025. RouteViews BGP Data. http://www.routeviews.org/. Accessed: 2025-01-21. Dates used: January 2025.

[22] Eric Wustrow, Colleen M. Swanson, and J. Alex Halderman. 2014. TapDance: end-to-middle anticensorship without flow blocking. In *Proceedings of the 23rd USENIX Conference on Security Symposium* (San Diego, CA) *(SEC'14)*. USENIX Association, USA, 159–174.

[23] Eric Wustrow, Scott Wolchok, Ian Goldberg, and J. Alex Halderman. 2011. Telex: anticensorship in the network infrastructure. In *Proceedings of the 20th USENIX Conference on Security* (San Francisco, CA) *(SEC'11)*. USENIX Association, USA, 30.

[24] Abdullah Yasin Nur and Mehmet Engin Tozal. 2018. Cross-AS (X-AS) Internet topology mapping. *Computer Networks* 132 (2018), 53–67. https://doi.org/10.1016/j.comnet.2018.01.011

A Supporting Tables

Table 1: Censorship statistics by country

Country	Domains	Censor. %	Res. AS	Uncens. AS
Iran	674,859	39%	57	400
China	638,232	23%	30	419
Turkmenistan	20,247	93%	2	310
Oman	6523	20%	1	354
Afghanistan	14,823	35%	2	323

Table 2: Top 10 ASes by path coverage for Iran

ASN	AS Name	Country
3257	GTT Communications Inc.	USA
174	Cogent Communications	USA
6453	TATA Communications (America) Inc	USA
12389	PJSC Rostelecom	Russia
6762	Telecom Italia Sparkle S.p.A.	Italy
49100	Pishgaman Toseeh Ertebatat Co.	Iran
1299	Arelion Sweden AB	Sweden
31713	Gateway Communications	Belgium
198154	Pars Abr Toseeh Ertebatat Co.	Iran
29049	Delta Telecom Ltd	Azerbaijan

Table 3: Top 10 ASes by path coverage for Iran with AS adjacency count

ASN	AS Name	AS Adjacency
3257	GTT Communications Inc.	4,529
174	Cogent Communications	14,209
6453	TATA Communications (America) Inc	7,037
12389	PJSC Rostelecom	1,932
6762	Telecom Italia Sparkle S.p.A.	7,116
49100	Pishgaman Toseeh Ertebatat Co.	44
1299	Arelion Sweden AB	8,260
31713	Gateway Communications	1,706
198154	Pars Abr Toseeh Ertebatat Co.	12
29049	Delta Telecom Ltd	319

Table 4: Number of top vantage ASes needed for 25%, 50%, and 75% coverage in each country

Country	25%	50%	75%
Oman	1	2	3
Turkmenistan	1	1	1
China	2	5	12
Iran	2	4	10
Afghanistan	1	2	5