Argus Lens: Innovating Internet Measurement Infrastructure via Relay Services

Zhiquan Wang
Tsinghua University, China
wzq22@mails.tsinghua.edu.cn

Changqing An
Tsinghua University, China
acq@cernet.edu.cn

Jilong Wang
Tsinghua University, China
wjl@cernet.edu.cn

***Abstract*—This paper proposes an innovative approach to enhance network measurement by leveraging relay services as a complement vantage points (VPs) to existing infrastructures. Traditional methods, such as Looking Glass and cloud services, often face limitations in cost, geographic distribution, and operational flexibility. Our proposed framework enables researchers to easily conduct multi VP measurement studies with commercial relay services. The framework automatically manages relay connections and executes measurement tasks, while polygraph tests ensure the integrity of measurement data against potential negative impacts from dishonest relay servers. Based on our system, benchmark tests on two popular Internet measurement topics, topology discovery and IP geolocation, demonstrate the framework's effectiveness, revealing the protocol compatibility, geographic distribution, and impact of relay services on measurement results. The findings show that relay-based measurement systems offer unique perspectives and capabilities compared to traditional platforms. Our work introduces novel tools and methods, as well as valuable datasets and practical experiences, which will inspire the development of community.**

***Index Terms*—measurement, relay, vantage point, VPN**

I. Introduction

Network measurement requires a wide distribution of vantage points (VPs) to collect data and derive conclusions. Biased or insufficient VP setups can compromise the accuracy of dataset, or even provide misleading results. When performing measurements in Internet routing [1], topology [2], accessibility [3] or cloud systems [4], researchers typically aggregate data from multiple VPs to capture fact on a large scale to provide comprehensive insight. However, building large-scale measurement systems faces practical challenges from high costs and obtaining necessary cooperation. Looking Glass (LG) partially fills the gap. As an existing infrastructure, it enables researchers to execute measurement from widely distributed VPs. However, the lack of standardized operational methods, limited probing frequency, inability to customize commands, and high usage costs make it difficult to conduct Internet-scale measurements with LG. Cloud service providers (such as AWS, Azure, and GCP) allow researchers to deploy VPs remotely and are widely accepted by the community. However, cloud service are not suitable for measurements that require a large number of VPs in different geographic locations because large cloud data centers are typically concentrated in a few economically advanced and geopolitically stable areas. Obtaining comprehensive VPs requires researchers to use multiple regional cloud providers, resulting in repetitive deployments, complex payment processes, and higher costs.

Fortunately, the flourishing relay service providers (such as NordVPN, ExpressVPN, ProtonVPN, etc.) have built global relay networks that aligns with the infrastructure needs of measurement research. These service providers often use terms such as *VPN* in brand names and marketing. Technically, VPN protocols enable comprehensive virtual network construction. The relay network we describe is merely a specific implementation of virtual networks and can be built using VPN technologies. Relay services encrypt users' traffic with dedicated protocols and forward them to the destination through relay servers, providing a solution for protecting privacy and security. Relay providers deploy connectable servers worldwide to minimize latency impacts and comply with regional policies. They offer superior flexibility compared to LG for full control over probing packets, while their geographical diversity compensates for the limited distribution of individual cloud providers. Relay services can serve as a powerful complement to existing measurement infrastructure, enhancing efficiency and providing more comprehensive fact.

This work proposes relay services as a novel supplement to existing measurement infrastructures while addressing practical implementation challenges. We designed and implemented an automated framework to assist researchers in quickly setting up measurement systems based on relay services. This framework can automatically manage connections of multiple relay servers and complete predefined measurement tasks through each relay server. Additionally, we designed a series of automated *polygraph* tests to check the properties of the relay servers, ensuring that the results are not compromised by dishonest relay servers. Based on this framework, we designed benchmarks for two popular topics to demonstrate the significant potential of relay networks in practice. We selected three large commercial and two free relay service providers to run the benchmarks. With the metadata collected by the framework, we analyzed the properties of these services that could impact Internet measurement. We examined the protocols they used, the geographic and Internet distribution of relay servers, and their cost. We demonstrated the compatibility of different protocols with measurement techniques and explored potential side effects. The results show that

This research is supported by Tsinghua University - Beijing Qihu Technology Co., Ltd Joint Research Center for Cyberspace Surveying and Mapping.

ISBN 978-3-903176-72-0 © 2025 IFIP

our framework is practical in real-world scenarios, and the measurement network based on relay services provides novel datasets that many other measurement platforms cannot offer.

Contribution. This work presents the following contributions: (1) We propose a methodology for utilizing relay services in Internet measurement and implement an automated relay-based framework. (2) We comprehensively analyze the mainstream relay service ecosystem, assess technical feasibility, and quantitatively evaluate how their properties impact measurement results. (3) We benchmark our relay-based measurement method against two popular topics, demonstrating its effectiveness. (4) All tools and datasets developed in this work are open-sourced[1], enabling researchers to readily apply our methods or evaluate relay servers' utility in future research.

II. Motivation and Background

High-quality measurement datasets rely on multiple VPs. Platforms like CAIDA's Ark [5] and RIPE Atlas [6] offer community-shared infrastructure, enabling researchers to perform tasks like Ping and Traceroute with extensive VP networks. Although these LGs lower measurement barriers, they come with limitations such as restricted rates, non-customizable commands, and scale-based costs, which impede Internet-scale measurements. Consequently, researchers often still need to build their infrastructures. However, it can be challenging particularly when VP diversity is necessary.

To address the challenges in constructing infrastructure, we have innovatively explored the feasibility of using commercial or free relay services. These services often offer a vast number (up to 6343) of relay servers at a low cost. As a primary feature touted by providers, these servers are distributed across various countries and cities worldwide, providing advantage for establishing a global measurement system with diverse network locations. Furthermore, the protocols used by mainstream providers to connect users to servers are compatible with most measurement techniques. Although this introduces additional variables that need to be understood and properly managed, such as extra hops or delays, and some technical challenges that need to be resolved, such as how to detect relay servers that are dishonest about geographical location [7], the potential for relay-based measurement system is immense. Researchers can enhance infrastructure with numbers of VPs on a modest budget. These insights enable researchers to collect comprehensive datasets, improving research accuracy and benefiting the measurement community's development.

Network location of VPs. Researchers often describe the network location of a VP with originated BGP prefix or AS. Different network locations usually imply different routing paths, AS relationships, policy constraints, etc., and often correlating with geographical locations. The diversity of VP locations in LG alleviates researchers' needs. However, the strict limitations in LG may hinder researchers from taking advantages of VP locations, making them impractical for specific contexts. Cloud-based VP typically locate in data centers, which are closer to the core of the Internet topology rather than the access ISP where edge users reside. Data center IP ranges may impose stricter application limitations like CAPTCHA challenges, affecting measurement outcomes. As shown in Figure 1, we tested the HTTP latency[2] to the Microsoft Bing website with three groups of VP: cloud VPS, edge user, and relay service. Result shows significantly lower latency in VPS group compared to others, suggesting location-dependent bias in measurement outcomes.

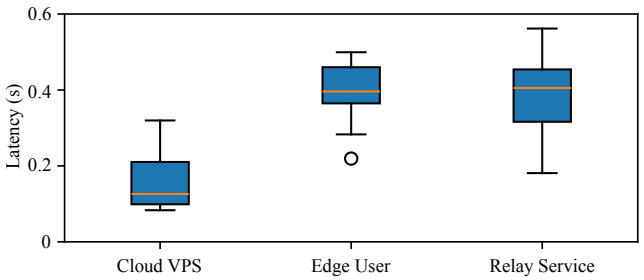

Fig. 1. Latency to Bing reveals cloud VPS in data centers have significantly lower latency compared to edge users and calibrated relay services.

Budget for purchasing services. In previous research, we used mini PCs, as shown in Figure 2, as portable measurement devices deployed within the network of partners. These low-power devices, costing over $100, are capable of executing tasks and transmitting results back. Cloud providers typically offer flexible subscription-based or pay-as-you-go VPS plans. The price for such products typically starts at $5 per month per host. Providers that support pay-as-you-go billing reduce the billing unit to minute level, with monthly fees still starting at a few to a dozen dollars. It is usually also necessary to pay for traffic or bandwidth. RIPE Atlas offers a credit-based payment plan, allowing researchers to earn credits through contributions such as providing VPs and then use these credits for measurements. Relay service providers typically offer a subscription payment model, where paying a few dollars per month grants access to all servers without any traffic limitations. Dedicated hardware has the highest unit price, but it offers the advantage of indefinitely use after a one-time purchase. Cloud services have a pricing advantage, with their flexible billing models making them attractive for both temporary and permanent measurement facilities. They incur significant costs when multiple VPs are required. RIPE Atlas is appealing to experienced researchers who have usually accumulated a large amount of credits. Relay services offer the lowest average price per VP, which makes them attractive as standalone or supplementary infrastructure.

Willingness to cooperate. Cloud service can be obtained through purchase. However, if VPs with different network locations are necessary, it typically requires collaboration with ISP to deploy hardware devices within their network. Such collaborations usually involve negotiations, yet ISPs

[1]https://github.com/flynnoct/argus-lens

[2]HTTP latency is time difference between sending an HTTP request and receiving response from Bing server. DNS were pre-cached to prevent impact.

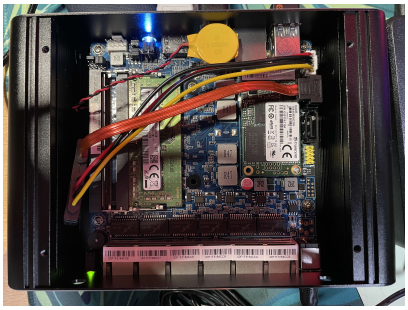

Fig. 2. A previously deployed probe device for collecting network measurement data, capable of running Linux, costing approximately $100.

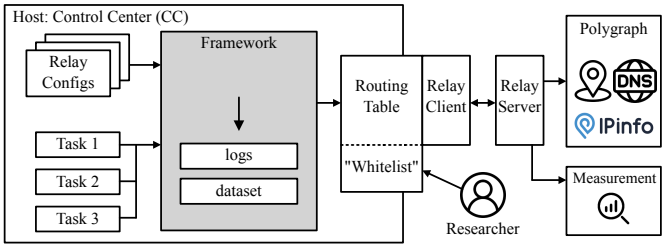

Fig. 3. Design of Relay-Based Measurement Framework

may refuse these proposals due to concerns over network security, privacy, and other reasons. This makes the practical devices deployment highly challenging, especially for large-scale deployments conducted globally.

To our knowledge, this is the first work to examine usage of relay networks as measurement infrastructure, analyze how their properties influence measurement results, and provide the community with automated tools. While some works have employed application-layer proxies in their methodologies, they either have not made their tools publicly available or have focused exclusively on specific measurement domains like DNS. These works have not conducted analysis on IP-level relay services capable of supporting more traffic types, such as traceroute, which we discuss in Section VII.

III. Framework Design

In this section, we introduce the design of our automated relay-based measurement framework and explain how it aids researchers in accurately and efficiently completing their measurement tasks. The design and workflow of the entire framework are illustrated in Figure3. The framework is designed to accept multiple relay service configurations as input. For each individual relay service configuration, the framework undergoes the following four phases described in this section and, upon completion, cycles to the next relay configuration until the measurement tasks are completed on all relay servers. In practice, we encountered four challenges. First, we discovered some relay servers with exaggerated claims, typically misrepresenting their geographical locations. Second, we experienced security lockouts designed to prevent abuse when frequently operating relay services. Third, these relay services are typically designed to establish only one connection at a time and route all traffic through it, making concurrent measurements difficult. Finally, the relay services themselves introduce errors that required calibration and correction. We addressed these challenges in our implementation.

Connection Initialization. The framework reads a relay configuration that includes commands for connecting and disconnecting from relay server. It calls the specific relay client app (e.g., OpenVPN) to run the connection command, establishing a connection to the relay server while creating a virtual network interface (VNI). It then configures the host's routing table to direct all or part of the traffic through the VNI. The relay client encrypts traffic entering this VNI with tunnel protocol and forwards it to the relay server. This process is application-independent since the OS's routing table controls the traffic routing rules. We made it possible to simultaneously create multiple VNIs and conduct parallel measurements through this mechanism: we grouped target IPs and employed a sophisticated module to control the matching between IP groups and VNIs through routing tables. After completing a measurement subtask, these mappings rotate, ensuring all IP groups executed all subtasks through all VNIs.

Relay Server Polygraph. Weinberg et al. [7] reported that proxy service providers might exaggerate the geographical location. To avoid measurement errors caused by unchecked servers, we designed a series of polygraph measures in the framework to detect brag behavior. The framework compares server's claimed location (at city level) provided by the researcher when configuring the connection. It converts location into latitude and longitude with the OpenStreetMap API [8], and pings the five IPs closest to the claimed location from a known IP landmark dataset, and reporting any abnormally results. To avoid misunderstandings, we also select a set of IPs located in globally significant locations, such as famous city for data centers (e.g., Singapore, Frankfurt and San Francisco), and perform pings. If abnormally low RTTs appear, we have evidence to believe that the relay server is actually closer to these significant locations, indicating that service providers have exaggerated the geographical location of the servers. Additionally, relay servers may use different ingress and egress IPs. Therefore, we access the IPInfo [9] API via the relay server, a third-party perspective, to test the egress IP of the server. Lastly, to record the network location of the relay server, we use traceroute to trace several common DNS server IPs and map the IPs in the path through BGP longest-prefix match to prefixes. The RouteViews Prefix to AS mappings [10] dataset provides mappings of IP Prefix to AS from global BGP routing tables. We look up the prefix mapped to a unique AS in this dataset. The BGPHE dataset provides inferred ASN to organizations or companies. The framework records the first ASN result that does not belong to the relay service provider itself. This usually represents the ISP providing WAN access to this relay server. Using multiple DNS IPs for tracing is to discover potential multi-WAN situations as much as possible. These ASNs are recorded and reported to the researcher as the

network location inspection result. All the above test results are saved in the framework's log files as reference information for the researcher to check result's credibility.

Running Measurement Tasks & Disconnecting. The framework reads a series of preset measurement commands, such as traceroute, scamper, or curl, executes them sequentially, and records results. This data constitutes the measurement dataset expected by researcher. After executing all measurement commands, the framework uses the disconnect command to terminate the connection with the relay server. Subsequently, the framework establishes a connection with another relay server and performs the measurement tasks until all relay servers provided in the collection are traversed.

In practice, researchers only need to provide the framework with the required relay configurations and the measurement tasks to be completed. The relay configuration includes metadata that will be used for the polygraph, such as provider ID and claimed geographical location, as well as the commands for connecting to and disconnecting from the relay server. These commands are usually very easy to obtain from service provider. This plug and play feature also reduces the cumbersome process researchers face when using cloud VPS systems, as it eliminates the need to configure the runtime environment for each VP individually, especially the repetitive tasks required to re-initialize instances after releasing pay-as-you-go instances. The measurement tasks include a series of desired measurement commands or scripts that framework will execute after establishing a connection with each relay server. All required input configurations are included in a structured configuration file, making it easy to set up the system.

We also identified some implicit limitations, such as certain relay servers unpredictably failing to forward Internet traffic. To address this in the polygraph process, we access IPInfo via the relay server to obtain the third-party reported IP, successfully retrieving the egress IP confirms successful Internet access through the server. We also noticed that some providers restrict how often you can connect to their relay servers. Initially, frequent connections to the same provider's relay servers (several times per minute) led to failures. To mitigate this, we developed a mechanism to rotate through different providers, ensuring servers from the same provider aren't used consecutively in a short time frame. If providers are insufficient, the framework adds time intervals to avoid restrictions. These measures effectively prevent connection failures through a scheduled connection sequence.

IV. Properties of Relay Services

In this section, we explore the properties of relay providers and protocols that might affect their application in Internet measurement, such as supported protocols, server distribution, and cost, etc. This exploration is based on data collected through polygraph described in Section III while running the benchmarks outlined in Section VI on five providers.

Provider	Nord	Express[1]	Proton	VPNBook	VPN Gate
WireGuard	△[2]		✓		
OpenVPN	✓	✓	✓	✓	✓
L2TP/IPsec					✓
IKEv2/IPsec	✓	⋆	⋆		
PPTP				✓	
SSTP					✓

[1] ExpressVPN also offers a private VPN protocol called Lightway, which is exclusively available within their client software.
[2] ✓: Supported and manually configurable.
⋆: Supported but **not** manually configurable.
△: Providers have implemented modifications based on this protocol.

TABLE I
Relay protocols supported by service providers.

A. Supported Protocols

Based on commonly observed facts, we categorize service providers' protocol support strategies into three types. First, necessary configurations are fully provided to allow users to set up relay connections on any devices (✓ in Table I). This open strategy is ideal for measurement as it allows us to programmatically control relay services. Second, service providers offer standard protocol connections, but the connection is controlled by dedicated client software, which may not be supported by all mainstream OSes (⋆ in Table I). Although this strategy does not alter the protocol compatibility, it may limit the measurement system. For instance, ProtonVPN's IKEv2/IPsec protocol is only supported on macOS, not on Windows or Linux. Hence, Linux-based measurement systems are unable to use IKEv2/IPsec. The last type involves service providers using completely private protocols or modified standard protocols. This strategy is the least favorable, as it requires us to analyze or test each unique protocol to determine whether it can carry the measurement traffic correctly.

As shown in table I, all relay providers offer OpenVPN as a connection method. This open-source VPN protocol implementation provides good flexibility, and its third-party client software covers nearly all mainstream OSes. This means that it can even be applied to measurements on mobile platforms, such as Apple iOS or Android. As an emerging high-performance tunneling protocol, WireGuard is gradually being accepted by service providers and users, serving as the successor to OpenVPN in the industry. ProtonVPN provides WireGuard guides, NordVPN claims their dedicated protocol *NordLynx* is based on WireGuard design, and although ExpressVPN does not offer a WireGuard connection, they have also engaged in discussions about this protocol [11]. While there are still three service providers offering IKEv2/IPsec as a connection method, the fact that they do not provide guides and can only be connected using their dedicated clients also indicates that they are gradually phasing out this protocol. Meanwhile, L2TP/IPsec, PPTP, and SSTP have been abandoned due to poor compatibility or proven vulnerabilities.

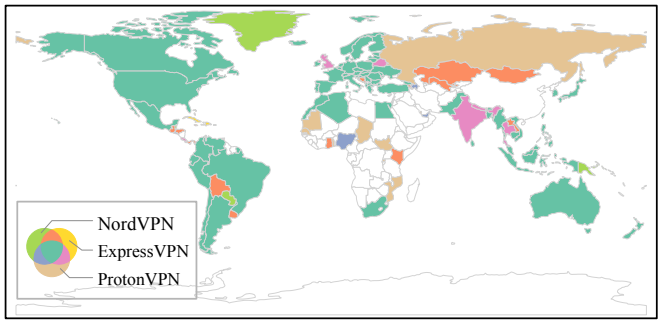

Fig. 4. Global servers distribution of relay providers, with the Venn diagram's color scheme indicating countries that host servers from multiple providers.

B. Server Distribution

As shown in Table II, each commercial relay service provider claims to offer more than 500 available servers, with NordVPN even boasting 6,343 distinct relay servers. Considering that multiple VPs located within the same BGP prefix or AS might not provide additional meaningful data in large-scale measurement, we analyzed the prefixes and ASes originating from the traffic egress IPs of relay services, while filtering out the unconnectable servers. We screened 483, 104, and 222 VPN servers located in different BGP prefixes for three commercial services, respectively. The benchmark data in Section VI is derived from these servers. These different BGP prefixes originated from dozens of different ASes, with the AS providing the most prefixes for NordVPN, AS136787 (owned by commercial company PacketHub), including 68.1% (329 of 483) of the prefixes, while ExpressVPN and ProtonVPN had 31.7% and 35.6% of their prefixes originating from AS206092 (IPXO) and AS212238 (DataCamp), respectively. These commercial companies primarily offer IT infrastructure and network code leasing services, indicating that commercial relay service providers choose to lease IPs and servers in bulk from third parties to provide relay services. In contrast, free relay service providers offer significantly fewer available VPs, and the free pricing model also makes it difficult for them to lease extensive network resources from other commercial companies. VPNBook has 85.7% of its prefixes originating from AS16276 (OVHCloud), a cloud service provider. Nobori et al. [12] have described how VPN Gate operates entirely through volunteers contributing their local networks, and we expect it to provide a broader distribution of prefixes. Actually, only 28.8% of the prefixes come from AS4766 Korea Telecom, with other ASes contributing a lower proportion of prefixes, indicating that the primary contributors to VPN Gate are residential ISPs, providing a different perspective from data centers for our network measurements.

C. Geographical Habitat

We used three data sources to cross-validate the geographical location of servers, as a single dataset is insufficient to provide accurate IP geolocation [14]. As shown in Table II, we counted the number of countries and regions where each relay service provider claims to offer relay servers. Commercial relay service providers typically offer relay services in about 100 regions, whereas free relay service providers can only offer relay services in fewer than ten regions. Figure 4 shows the geographical distribution of three large commercial relay providers. According to our connection tests of relay servers, we found that the number of countries that NordVPN and ProtonVPN could connect to was comparable to their claims, providing 96.4% and 89.0% availability respectively, while ExpressVPN performed worse, with the number of successfully connected regions being only 62.9% of its claimed figure. IPInfo provided a perspective from a third-party IP geolocation database, with query results significantly lower than the number claimed by the provider service, and the maximum discrepancy is 22 regions for NordVPN. We tested latency from relay servers to IP landmarks and we observed some geographical mismatches, such as NordVPN's PE1 server, which claims to be located in Peru. RTT test from our VM to PE1 was approximately 169ms. 148.163.223.233 (referred to Atlas-CL) is a RIPE Atlas probe located in Santiago, Chile, which we use as an IP landmark. The RTT from our VM to Atlas-CL is only 171ms (via the PE1 relay). With this 2ms latency difference, we estimate that the geographical distance between PE1 and Atlas-CL should not exceed 200km. However, the actual geographical distance between the two is over 2450km. Conversely, we also conducted RTT tests from third-party LG probes claimed to be located in Santiago to PE1, with the lowest RTT being only 0.913 ms. Therefore, we have reason to believe that the NordVPN server claimed to be in Peru is not located within a few hundred kilometers of Santiago, which means the NordVPN server claimed to be in Peru is not actually within Peru. The tunnel remote endpoint of ExpressVPN provided a lower Ping response rate, making it difficult to estimate the tunnel RTT. Therefore, we excluded ExpressVPN server data from this test and subsequent geolocation benchmarks. We tested each server located within the prefix using a relatively generous range (three times the geographical distance estimated based on RTT). The other two commercial relays had an error rate of less than 9%, and we excluded servers with potentially incorrect geographical locations from subsequent benchmarks. We also found that this phenomenon mostly occurred in less developed regions, possibly due to the lack of good data center infrastructure in these areas, leading to the use of nearby servers as substitutes.

D. Cost Estimation

The intention of using relay servers as an extension of the existing measurement infrastructure was to enhance the diversity of VPs with a relatively low budget. We evaluated the costs of purchasing relay services in Table II. Most commercial providers use simple unlimited subscription plans, where traffic is not limited after payment for a certain period. In contrast, cloud services employ more complex pricing models, typically combining periodic subscription fees for hosts with separate traffic charges. Some providers offer flexible pay-as-you-go options, allowing payment based on actual usage

Provider	Claimed Server	Available Region (claimed)	IPInfo[1]Region	BGP Prefix	AS	Allowed Device	Price
NordVPN	6343	107(111)	85	483	25	10	$12.99/mo
ExpressVPN	500+	66(105)	55	104	21	8	$12.95/mo
ProtonVPN	4950	81(91)	73	222	33	10	$9.99/mo
VPNBook	12	6(6)	6	7	2	∞	free
VPN Gate	-[2]	5(-)[2]	5	52	24	∞	free

[1] IPInfo is a third-party database providing IP address information. We collected country or region for IPs from IPInfo as a reference, despite its potential inaccuracy. [13]
[2] VPN Gate is entirely operated by volunteers, offering no guarantees on regional or server count.

TABLE II
SERVER AND NETWORK DISTRIBUTION OF SERVICE PROVIDERS.

without monthly commitments for short-term use. RIPE Atlas provides measurement services using platform credits. Each ping measurement from a single VP requires 10 credits, while a traceroute requires 20. The most accessible way to obtain credits is by sponsoring a probe, which earns 21,600 credits daily when online. For our benchmark in Section VI, the measurement runtime using relay services is approximately 4 days. With the typical minimum subscription period of 1 month, the maximum cost for a single commercial service provider is $12.99. For cloud services, ensuring at least one VP per region's data center while selecting the lowest performance VPs to optimize costs, an experienced researcher can complete system initialization and benchmarking within 48 hours. The costs amount to $6.71 for Amazon AWS's 29 VPs, $16.08 for Microsoft Azure's 50 VPs, and $69.33 for Google Cloud Platform's 31 VPs[3]. For RIPE Atlas, completing the experiment requires more than 5 million credits — equivalent to a single probe being continuously online for over 200 days. While experienced researchers might afford measurements using accumulated RIPE Atlas credits, and cloud services offer flexible payment options for short-term tasks, relay services provide significantly broader geographic VP coverage than mainstream cloud providers at relatively low cost. This eliminates the complexity of managing multiple cloud service providers to expand coverage, substantially lowering measurement barriers. Therefore, using relay services for network measurements remains highly attractive.

V. IMPACT ON MEASUREMENT

In this section, we analyze the impact of using relay servers on the execution of traceroute and latency measurements. It is impossible to cover all measurement contexts in this paper, so the impact on other types of measurements needs to be discussed independently. In addition to encrypting and encapsulating data transmitted through the tunnel, the relay server functions as a gateway. It performs two operations on the data packets: Source NAT and forwarding, preserving other fields in the IP packet header unchanged except for the source IP and port. It also provides the capability to fully forward IP layer response packets, including both IP and ICMP packets, back to source host. Therefore, a VPN tunnel can carry any probe packets at IP layer and above.

[3]This estimate excludes traffic fees, as traceroute and ping requirements are minimal (under 1GB per VP).

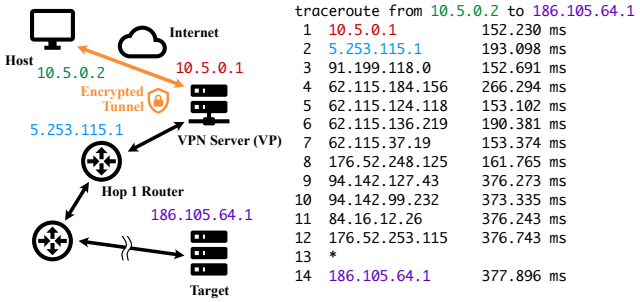

Fig. 5. An example demonstrates the packet path of traceroute through tunnel.

Traceroute requires a series of data packets with incrementally increasing TTL values in the IP header to trigger ICMP Time Exceeded responses from intermediate routers. It also requires ICMP forwarding capabilities to return the ICMP response packets, which contain the router's IP (the exact information we aim to obtain through measurement), to the host. Some works use different upper-layer protocols, such as tcptraceroute [15] to bypass firewalls and other middle-box devices, or Paris Traceroute which deliberately controls IP packet port and other stream identifiers to avoid the effects of load balancing devices that cause fake IP links. These variations of traceroute improve measurement accuracy and are widely accepted in the industry. And these variations do not fundamentally change the principles of traceroute measurement, therefore, they do not alter compatibility.

When running Traceroute, although the VPN tunnel is established over the Internet, spanning multiple hops, the entire tunnel is perceived as a single hop directly connecting client's and server's interfaces for internal packets. For Traceroute measurements, the VPN tunnel appears as the first hop which is easy to identify from the entire traceroute result because virtual connections within the tunnel typically use private IPs. Figure 5 shows the results of a UDP Paris Traceroute conducted through a tunnel. The server assigned VNI of our VPS the private IP 10.5.0.2, with the endpoint IP being 10.5.0.1. The server also used this virtual LAN IP to respond to Traceroute requests. In other words, if we consider the server as a VP, the first hop detected in the Traceroute path is 5.253.115.1, and so on. Therefore, handling the extra hop introduced by Traceroute is relatively straightforward. We

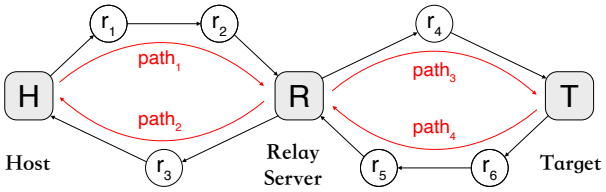

Fig. 6. Packet paths via relay server in a network with asymmetric routing.

simply need to remove the private IP hop at the beginning and consider it as the VPN server's WAN IP.

The method for measuring latency typically involves timing from the moment a probe packet is sent until a response packet is received, thereby obtaining the Round Trip Time (RTT) for the link or service. This RTT includes the link delay from the measurement host to the target host, the processing delay at the target host, and the return delay of the response packet to the measurement host. This type of latency measurement is not limited to using the ICMP Ping tool to measure link delay; it can also be extended to other areas of measurement. For example, the performance of website services can be assessed through HTTP response delays [16], or the status of DNS services can be evaluated through DNS response delays [17].

When using ping to measure the link delay, assume the forward path is $[\mathrm{A}-\mathrm{B}-\ldots-\mathrm{C}]$. Although using the result of $\mathrm{RTT}_{\mathrm{A-C}}-\mathrm{RTT}_{\mathrm{A-B}}$ to estimate $\mathrm{RTT}_{\mathrm{B-C}}$ is not entirely accurate, it can still provide researchers with some evidence to evaluate performance. The accuracy of this estimate may be influenced by various factors, including but not limited to the processing delay at host B, routing asymmetry, and possible route changes during the measurement. However, when the intermediate server (host B) is a relay server, host B serves as an essential anchor for the return path, as the destination of the response packet is host B, which then forwards it to the original host. This ensures that $\mathrm{RTT}_{\mathrm{A-B}}$ and $\mathrm{RTT}_{\mathrm{B-C}}$ are independent of each other, eliminating the impact of routing asymmetry.

Under the network topology shown in Figure 6, we analyze the method of measuring end-to-end RTT using a relay server. Let t_{p1}, t_{p2}, t_{p3}, and t_{p4} denote the transmission delay over each segment of the path, and let t_{tun} represent the processing delay introduced by the relay protocol. In the experiment, we deployed two cloud servers located in the United States and Singapore and continuously measured several RTT values: $\mathrm{RTT}_{\mathrm{total}}$ $(= t_{\mathrm{p1}} + t_{\mathrm{p2}} + t_{\mathrm{p3}} + t_{\mathrm{p4}} + t_{\mathrm{tun}})$, $\mathrm{RTT}_{\mathrm{relay}}$ $(= t_{\mathrm{p1}} + t_{\mathrm{p2}} + t_{\mathrm{tun}})$, $\mathrm{RTT}_{\mathrm{direct}}$ $(= t_{\mathrm{p1}} + t_{\mathrm{p2}})$ and $\mathrm{RTT}_{\mathrm{real}}$ $(= t_{\mathrm{p3}} + t_{\mathrm{p4}})$. Figure 7 illustrates a comparison between $\mathrm{RTT}_{\mathrm{relay}} + \mathrm{RTT}_{\mathrm{real}}$ and $\mathrm{RTT}_{\mathrm{total}}$.

The measurement results indicate that $\mathrm{RTT}_{\mathrm{total}} - \mathrm{RTT}_{\mathrm{relay}}$ can effectively estimate $\mathrm{RTT}_{\mathrm{real}}$, as both exhibit consistent mean values. We employed the Augmented Dicky-Fuller (ADF) Test to verify whether the difference between the two sequences is stationary, and subsequently used the Two One-

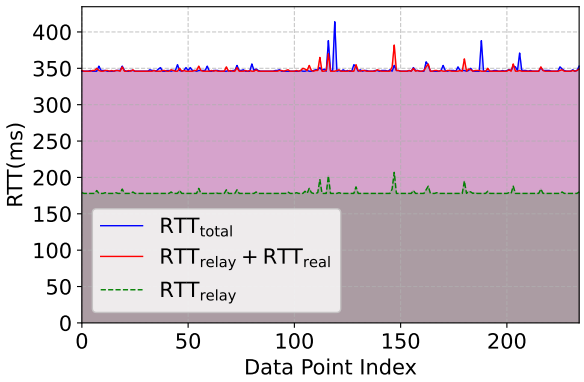

Fig. 7. Comparison of $\mathrm{RTT}_{\mathrm{total}}$, $\mathrm{RTT}_{\mathrm{relay}}$, and $\mathrm{RTT}_{\mathrm{real}}$.

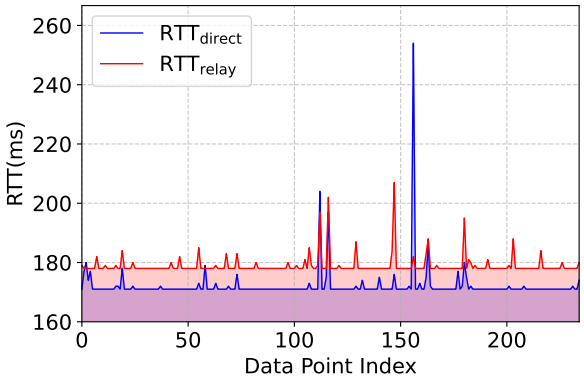

Fig. 8. Comparison between tunnel latency and WAN latency between host and relay server.

Sided Tests (TOST) method to examine whether there is a significant mean difference between them. The p-values of the hypothesis tests are all much less than 0.01, indicating that the difference sequence is stationary and the mean difference lies within ± 2.1 ms (approximately 0.2%), with an average error of 0.75 ms. These results demonstrate the effectiveness of this estimation method.

In addition, we measured the latency introduced by the relay protocol itself. This was achieved by comparing $\mathrm{RTT}_{\mathrm{direct}}$ when the measurement host pings the relay server directly with the $\mathrm{RTT}_{\mathrm{relay}}$ when via the tunnel, as shown in Figure 8. We also used the ADF Test to examine the stability between the two, and the results indicate that, on our testing platform, tunnel operation brings a stable temporal overhead of approximately 6.77 ms. This result can serve as a baseline reference for evaluating relay performance under different protocols and configurations.

In actual measurements, delay may vary due to non-relay factors such as network congestion and routing fluctuations. Therefore, statistical methods are required for error correction. In addition, during measurement, the additional delay introduced by the relay should be reasonably adjusted according to the actual number of round trips, and appropriate statis-

Name	Geolocation		Uplink ASN	
t-uswest	US	San Francisco	AS1299	Arelion
t-useast	US	Ashburn	AS174	Cogent
t-fra	DE	Frankfurt	AS1299	Arelion
t-sg	SG	Singapore	AS7473	Singtel
t-sp	BR	São Paulo	AS3356	Level 3
t-hk	HK	Hong Kong	AS3491	PCCW Global
t-sh	CN	Shanghai	AS4812	China Telecom

TABLE III
GEOLOCATION AND UPLINK AS INFORMATION OF VM-TS.

System		VPN Services	VM-Ts	RIPE Atlas
IP addr	found	38,600	9,474	56,581
	VPN only[1]	-	30,616	29,319
IP link	found	124,573	12,675	81,398
	VPN only	-	115,945	115,823

[1] The 'VPN only' row indicates the number found exclusively by the relay and not by others.

TABLE IV
IP ADDRESSES AND LINKS DISCOVERED BY MEASUREMENT SYSTEM.

# of IP	ASN	Organization	CC
2893	AS13335	Cloudflare, Inc.	US
1558	AS8075	Microsoft Corporation	US
1148	AS4134	China Telecom Backbone	CN
1108	AS15169	Google LLC	US
1031	AS3356	Level 3 Parent, LLC	US

TABLE V
TOP 5 ASES WITH HIGHEST NUMBER OF IPS DISCOVERED BY RELAY SERVERS BUT NOT DETECTED BY RIPE ATLAS.

tical indicators should be selected (for example, using the minimum delay for geolocation estimation). Our experiments demonstrate that it is feasible to correct the systematic error of RTT measurement by measuring ping delay inside the tunnel, and this automatic monitoring function has already been implemented within the framework.

VI. BENCHMARK

We conducted a benchmark using two classic topics in Internet measurement to assess the effectiveness of using relay services for measurements. The results demonstrate that relay services provide views that other measurement platforms, such as RIPE Atlas, cannot offer. They revealed a considerable number of unique measurement insights, showcasing the significant potential of relay services as infrastructure.

Testbed. Control Center (VM-CC) is a virtual machine used for scheduling management and dataset storage. *Telescopes* (VM-Ts) consist of 7 VMs used for measurement. Basic information of VM-Ts is shown in Table III. VM-Ts provide us with a baseline dataset for benchmarks.

A. Topology discovery via Traceroute

To enhance network topology inference through diverse IP link discovery, we employed Traceroute data from multiple VPs. We targeted a set of router IPs previously verified as reachable and responsive to Traceroute probes. Using VM-CC connected to various relay providers' servers, we executed UDP Paris Traceroute [18] on these targets with scamper [19]. Simultaneously, we conducted identical tests with RIPE Atlas and VM-Ts, with RIPE Atlas aggregating results across its VPs. For router interconnection analysis, we decomposed IP paths into IP link pairs—adjacent IPs within each path. To ensure measurement accuracy, we filtered out any IP link pairs containing BOGON IPs [20], as these reserved Special-Purpose IPs, when improperly used on inter-router interfaces, can lead to erroneous IP link assessments.

We conducted measurements on 929 non-anycast IPs from known responsive routers and extracted IP link pairs. This accurate target selection reduces the issue of non-unique targets caused by CDNs when using popular web datasets, thereby providing a uniform comparison. This benchmark demonstrates the system's ability to uncover connections between routers in topology discovery. The measurement results, as shown in Table IV, indicate that VPN system observed 38,600 unique router IPs and 124,573 unique IP links. In contrast, RIPE Atlas observed 56,581 unique IPs and 81,398 unique IP link pairs. Commercial relay services performed significantly better than free services, with NordVPN alone discovering over 33,000 IP and over 100,000 unique IP links. This observation aligns with our expectations — the commercial model enables more resources to set up VPN servers distributed globally and across various network locations, facilitating the observation of more extensive topology information.

We compared the VPN dataset with result from RIPE Atlas. There are 29,319 IPs and 115,823 IP link pairs were observed only in the VPN, accounting for 76.0%. We mapped these VPN-only IPs to AS with the Hurricane Electric dataset [21]. The top five AS with the most IPs and their corresponding organization names are listed in Table V. These mainly include Internet content providers and regional ISPs, and these IPs are typically located at the end of the IP path. We believe the primary reason for this is that relay services offer a unique VP perspective, and the paths into the end AS differ from those observed by Atlas probes, thus revealing more router IPs.

B. RTT-Based Geolocation Estimation

RTT has been a crucial reference for geographical distance estimation in IP geolocation works [22]–[25]. We used RTT measurements from multiple VPs to infer target IP host locations. For ground truth, we selected 935 landmarks (IPs from RIPE Atlas probes with known locations) and collected RTT data by pinging these landmarks using relay servers, RIPE Atlas, and VM-Ts. We also incorporated GeoLite2 [26] database as a baseline. To mitigate temporary network congestion or routing changes effects, we gathered latency data over several weeks. We applied Constraint-Based Geolocation (CBG) to establish a linear relationship between delay and distance, assuming optical fiber signal propagation at 200 km/ms [7]. Unlike RIPE Atlas and VM-Ts hosts, relay servers only forward probe packets, incorporating forwarding time in the total RTT. We estimated RTT from relay servers to landmarks by calculating the difference between total RTT and

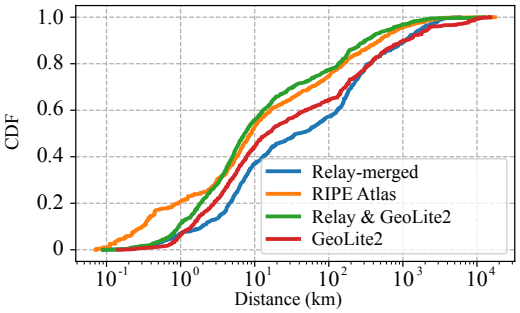

Fig. 9. Distance error CDF between RTT-based IP geolocation and ground truth. Adding VPN to GeoLite2 improved accuracy by 13.7% at 143 km.

tunnel RTT. Comparing our inferences with ground truth, we calculated distance differences between actual and estimated locations for each target IP, as illustrated in Figure 9. While relay-based inferences did not exceed RIPE Atlas or GeoLite2 accuracy, they demonstrated competitive performance, even outperforming GeoLite2 beyond the 200km range.

In the inference process of IP geolocation, a latency difference of 1ms typically signifies a range of 100km. This implies that improving localization accuracy below 100km necessitates data from multiple VPs located in the same city but in different positions. However, relay service providers have yet to furnish localization information beyond the city level (such as the precise geographical locations of data centers or servers). Consequently, we can only use the central positions of cities as estimates for VP locations, posing challenges for us to further enhance localization accuracy. Simultaneously, this may explain the observed rapid increase in latency data from VPNs beyond the 100km threshold. Moreover, professional IP geolocation databases may incorporate additional data sources such as reverse DNS [27]. Manaf et al. [22] previously discovered some country-level data inaccuracies in free geolocation databases, which we speculate might stem from other sources of inference data. Therefore, we attempted to merge the VPN latency data with the GeoLite2. The results show that the VPN latency data corrected numerous city and country errors in GeoLite2. This improved the geolocation accuracy within a range of 134 km by 13.7%.

In summary, due to their extensive server distribution, relay servers typically facilitate providing a lower bound guarantee for IP geolocation inference, a.k.a., preventing continent-level localization errors. When employing naive methods, relay services can offer additional performance enhancements for IP geolocation, thus demonstrating potential that makes us optimistic about their practical applications.

VII. Related Work

Previous works in Internet measurement have sporadically utilized relay networks such as VPN and proxy services, though primarily as supplementary data collection mechanisms rather than as comprehensive methodological approaches. Several works have employed VPN services for specific measurement purposes. Niaki et al. [28] conducted long-term measurements of Internet censorship behavior via ICLab, and Kashaf et al. [29] analyzed third-party infrastructure dependencies of African websites. However, neither study disclosed their complete measurement system design or technical details regarding relay-based measurements. Similarly, multiple works [30]–[32] used BrightData (formerly Luminati) proxies to investigate DNS ecosystem, with Chhabra et al. [30] providing valuable insights on latency and DNSSEC verification anomalies via proxies. While these works demonstrate the utility of relay services, they predominantly focus on application-layer proxies (HTTP/SOCKS) or restrict scope to DNS measurements, without exploring IP-level relay capable of supporting more diverse traffic patterns such as traceroute.

Khan et al. [33] measured the commercial VPN ecosystem, analyzing security, privacy, and server geographical distribution to establish a macroscopic understanding of VPNs. Nobori et al. [12] developed VPN Gate, a volunteer-based VPN service network that, while still operational, primarily relies on servers from a few academic institutions, limiting its effectiveness for diverse measurement research. Maghsoudlou et al. [34] examined VPNs during COVID-19, focusing on encryption, security, and privacy aspects. Unlike these studies that analyze the VPN ecosystem itself, our work explores VPNs' potential as measurement tools. By leveraging their geographical distribution and protocol compatibility advantages, we utilize the VPN ecosystem to discover diverse Internet properties, benefiting the broader research community.

LG provides an observation platform for the Internet. RIPE Atlas compiles and disseminates measurement data, facilitating result reuse and minimizing redundant measurement traffic [35]. Zhuang et al. [36] used crawlers and search engines to find hidden LG sites, while Periscope [6] consolidated interfaces across LG sites to improve efficiency. Our work achieves objectives similar to LG but employs distinct technical approaches. Specifically, our relay-based measurement system overcomes several LG limitations by providing researchers easy access to numerous VPs for conducting various compatible measurements.

VIII. Ethics and Privacy

Our approach is effective in practice, but its correct usage requires caution. Relay-based measurement methods can raise ethical, privacy, and abuse concerns. First, using relay services, especially those based on residential networks, for censorship measurements may expose contributors to unnecessary risks. Additionally, malicious or free relay servers could intercept or analyze traffic, posing threats to sensitive measurements. Finally, when using relay servers, adhere to fair use principles and service agreements, avoiding excessive traffic or resource consumption that may impact other users.

IX. Conclusion

This paper has introduced a novel approach to Internet measurement by leveraging commercial and free relay services as supplementary VPs to existing infrastructures. Our automated framework, equipped with polygraph tests to ensure data

integrity, demonstrates the practicality and effectiveness of using relay services in network measurement tasks. The benchmark tests on topology discovery and IP geolocation have revealed that relay-based measurements can provide unique and valuable insights, enhancing the scope and accuracy of Internet research. The results underline the significant potential of relay services to complement traditional platforms, offering broader geographic distribution, cost efficiency, and operational flexibility. By open-sourcing our tools and datasets, we aim to foster further innovation and collaboration within the research community, paving the way for more comprehensive and diverse Internet measurement studies.

ACKNOWLEDGMENT

This research is supported by Tsinghua University - Beijing Qihu Technology Co., Ltd Joint Research Center for Cyberspace Surveying and Mapping.

REFERENCES

[1] R. Anwar, H. Niaz, D. Choffnes, Í. Cunha, P. Gill, and E. Katz-Bassett, "Investigating interdomain routing policies in the wild," in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 71–77.

[2] Y. Shavitt and U. Weinsberg, "Quantifying the importance of vantage points distribution in internet topology measurements," in *IEEE INFOCOM 2009*. IEEE, 2009, pp. 792–800.

[3] G. Wan, L. Izhikevich, D. Adrian, K. Yoshioka, R. Holz, C. Rossow, and Z. Durumeric, "On the origin of scanning: The impact of location on internet-wide scans," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 662–679.

[4] R. K. Mok, H. Zou, R. Yang, T. Koch, E. Katz-Bassett, and K. C. Claffy, "Measuring the network performance of google cloud platform," in *Proceedings of the 21st ACM Internet Measurement Conference*, 2021, pp. 54–61.

[5] CAIDA. (2024, April) Archipelago (ark) measurement infrastructure. [Online]. Available: https://www.caida.org/projects/ark/

[6] V. Giotsas, A. Dhamdhere, and K. C. Claffy, "Periscope: Unifying looking glass querying," in *Passive and Active Measurement: 17th International Conference, PAM 2016, Heraklion, Greece, March 31-April 1, 2016. Proceedings 17*. Springer, 2016, pp. 177–189.

[7] Z. Weinberg, S. Cho, N. Christin, V. Sekar, and P. Gill, "How to catch when proxies lie: Verifying the physical locations of network proxies with active geolocation," in *Proceedings of the Internet Measurement Conference 2018*, 2018, pp. 203–217.

[8] F. Ramm, J. Topf, and S. Chilton, "Openstreetmap," *Die freie Weltkarte nutzen und mitgestalten*, vol. 3, pp. 978–3 865 413 758, 2010.

[9] IPinfo. (2024, April) The trusted source for ip address data. [Online]. Available: https://ipinfo.io/

[10] "RouteViews Prefix to AS mappings," https://catalog.caida.org/dataset/routeviews_prefix2as, dates used: April 2024. Accessed: April 2024.

[11] ExpressVPN. (2023, Jan) Will expressvpn use the wireguard protocol? [Online]. Available: https://www.expressvpn.com/blog/expressvpn-wireguard-update/

[12] D. Nobori and Y. Shinjo, "Vpn gate: A volunteer-organized public vpn relay system with blocking resistance for bypassing government censorship firewalls," in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 229–241.

[13] D. Komosny, M. Voznak, and S. U. Rehman, "Location accuracy of commercial ip address geolocation databases," *Information technology and control*, vol. 46, no. 3, pp. 333–344, 2017.

[14] O. Darwich, H. Rimlinger, M. Dreyfus, M. Gouel, and K. Vermeulen, "Replication: Towards a publicly available internet scale ip geolocation dataset," in *Proceedings of the 2023 ACM on Internet Measurement Conference*, 2023, pp. 1–15.

[15] M. Luckie, Y. Hyun, and B. Huffaker, "Traceroute probe method and forward ip path inference," in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, 2008, pp. 311–324.

[16] N. Naik and P. Jenkins, "Web protocols and challenges of web latency in the web of things," in *2016 Eighth International Conference on ubiquitous and future networks (ICUFN)*. IEEE, 2016, pp. 845–850.

[17] A. Hounsel, P. Schmitt, K. Borgolte, and N. Feamster, "Can encrypted dns be fast?" in *Passive and Active Measurement: 22nd International Conference, PAM 2021, Virtual Event, March 29–April 1, 2021, Proceedings 22*. Springer, 2021, pp. 444–459.

[18] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding traceroute anomalies with paris traceroute," in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, 2006, pp. 153–158.

[19] M. Luckie, "Scamper: a scalable and extensible packet prober for active measurement of the internet," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 2010, pp. 239–245.

[20] M. Cotton, L. Vegoda, R. Bonica, and B. Haberman, "Special-purpose ip address registries," Tech. Rep., 2013.

[21] H. Electric. (2024, April) Hurricane electric: Internet backbone and colocation provider. [Online]. Available: https://www.he.net/

[22] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos, "A look at router geolocation in public and commercial databases," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 463–469.

[23] Z. Hu, J. Heidemann, and Y. Pradkin, "Towards geolocation of millions of ip addresses," in *Proceedings of the 2012 Internet Measurement Conference*, 2012, pp. 123–130.

[24] B. Du, M. Candela, B. Huffaker, A. C. Snoeren, and K. Claffy, "Ripe ipmap active geolocation: Mechanism and performance evaluation," *ACM SIGCOMM Computer Communication Review*, vol. 50, no. 2, pp. 3–10, 2020.

[25] Z. Dong, R. D. Perera, R. Chandramouli, and K. Subbalakshmi, "Network measurement based modeling and optimization for ip geolocation," *Computer Networks*, vol. 56, no. 1, pp. 85–98, 2012.

[26] MaxMind. (2024, April) Geoip and geolite databases and web services. [Online]. Available: https://dev.maxmind.com/geoip

[27] O. Dan, V. Parikh, and B. D. Davison, "Ip geolocation through reverse dns," *ACM Transactions on Internet Technology (TOIT)*, vol. 22, no. 1, pp. 1–29, 2021.

[28] A. A. Niaki, S. Cho, Z. Weinberg, N. P. Hoang, A. Razaghpanah, N. Christin, and P. Gill, "Iclab: A global, longitudinal internet censorship measurement platform," in *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020, pp. 135–151.

[29] A. Kashaf, J. Dou, M. Belova, M. Apostolaki, Y. Agarwal, and V. Sekar, "A first look at third-party service dependencies of web services in africa," in *International Conference on Passive and Active Network Measurement*. Springer, 2023, pp. 595–622.

[30] R. Chhabra, P. Murley, D. Kumar, M. Bailey, and G. Wang, "Measuring dns-over-https performance around the world," in *Proceedings of the 21st ACM Internet Measurement Conference*, 2021, pp. 351–365.

[31] T. Chung, R. van Rijswijk-Deij, B. Chandrasekaran, D. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson, "A longitudinal, end-to-end view of the dnssec ecosystem," in *26th USENIX Security Symposium (USENIX Security 17)*, 2017, pp. 1307–1322.

[32] P. Bhowmick, M. I. Ashiq, C. Deccio, and T. Chung, "Ttl violation of dns resolvers in the wild," in *International Conference on Passive and Active Network Measurement*. Springer, 2023, pp. 550–563.

[33] M. T. Khan, J. DeBlasio, G. M. Voelker, A. C. Snoeren, C. Kanich, and N. Vallina-Rodriguez, "An empirical analysis of the commercial vpn ecosystem," in *Proceedings of the Internet Measurement Conference 2018*, 2018, pp. 443–456.

[34] A. Maghsoudlou, L. Vermeulen, I. Poese, and O. Gasser, "Characterizing the vpn ecosystem in the wild," in *International Conference on Passive and Active Network Measurement*. Springer, 2023, pp. 18–45.

[35] P. Gigis. (2017, Jun) Announcing daily ripe atlas data archives. [Online]. Available: https://labs.ripe.net/author/petros_gigis/announcing-daily-ripe-atlas-data-archives/

[36] S. Zhuang, J. H. Wang, J. Wang, Z. Pan, T. Wu, F. Li, and Z. Zhang, "Discovering obscure looking glass sites on the web to facilitate internet measurement research," in *Proceedings of the 17th International Conference on Emerging Networking EXperiments and Technologies*, 2021, pp. 426–439.