Server, Client, or Relay?
Dual-Role Detection of Circumvention Relays

Sultan Almutairi
North Carolina State University
Raleigh, North Carolina, USA
ssalmuta@ncsu.edu

Khaled Harfoush
North Carolina State University
Raleigh, North Carolina, USA
kaharfou@ncsu.edu

Yannis Viniotis
North Carolina State University
Raleigh, North Carolina, USA
candice@ncsu.edu

Abstract

In the ongoing cat-and-mouse game of Internet censorship, circumvention designs increasingly rely on obfuscation techniques, yet many still depend on a single IP address architecture. In this paper, we examine whether this design introduces an observable vulnerability. We hypothesize that these relays exhibit a dual-role behavior, acting both as servers that receive client connections and as clients that initiate new outbound connections. We ask whether this dual-role behavior can serve as a distinguishing feature for identifying relays, even when their traffic is fully obfuscated. To address the research question, we develop a three-stage detection pipeline and validate its core behavioral heuristic on a 17 TB dataset from the WIDE Project. We compare the behavior of a ground-truth set of relays (OpenVPN, WireGuard, SOCKS) against the general benign TLS traffic. The dual-role behavior provides successful classification, correctly identifying 23% of known relays while maintaining a negligible 0.18% False Positive rate against benign traffic, thereby validating the model's operational feasibility.

Keywords

Censorship, Circumvention, Detection

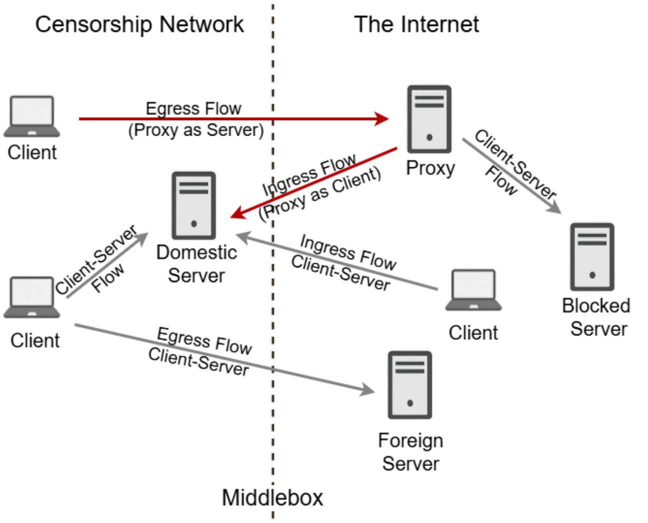

Figure 1: Dual-Role Behavioral Fingerprint. The diagram illustrates a foreign Proxy IP acting simultaneously as a Server (receiving client flow) and a Client (initiating relayed traffic).

1 Introduction

Internet censorship has become increasingly pervasive, restricting access to information and communication. Censors employ Deep Packet Inspection (DPI) to filter traffic based on blacklists of IP addresses and domains by inspecting DNS queries, TLS SNI fields, and HTTP host headers or keywords. These filtering techniques take several forms, including packet dropping, DNS poisoning, and the injection of TCP reset messages to disrupt access to censored content [7]. To bypass these restrictions, circumvention designs rely on external proxies that relay user requests outside the censored network. The designs typically add an additional layer of defense above the standard proxy protocol to hide their activity.

The first defense is to obfuscate the link between client and proxy. In general, obfuscation takes two forms: mimicking allowed traffic or encrypting traffic so that it appears indistinguishable from random noise [3, 9]. Both approaches make proxy traffic blend with benign background traffic, making it harder for censors to detect and block without causing collateral damage by interfering with benign servers. In response, censors advance their DPI middleboxes and use active probing, sending crafted packets to suspect servers to identify and block proxies. Circumvention designs then adopt a second defense with a probe-resistant technique, where a proxy remains silent to unsolicited probes unless the client authenticates [2]. This ongoing dynamic represents what is widely known in the literature as the cat-and-mouse game.

Despite these defenses, obfuscation and probe resistance protect only the link between client and proxy. They do not hide the broader traffic patterns that the proxy generates. Prior work shows that state-level censors possess extensive monitoring capability across their networks [7] and even over transit traffic [5], enabled by middleboxes that can be triggered by outbound, inbound, and transit flows that observe domains and act upon blacklists at scale. Such capability allows the censor to infer the operational role of each server within its monitored space.

Our hypothesis is that a relay server exhibits a distinct dual-role behavior that distinguishes it from benign servers. Normal servers primarily act as servers—they receive inbound connections from clients. Normal clients, such as user devices, primarily act as clients, which initiate outbound connections to other servers. A relay, by contrast, does both: 1) It acts as a server by receiving inbound

Free and Open Communications on the Internet 2026(1), 58–61

connections from clients seeking circumvention and 2) It simultaneously acts as a client by initiating new outbound connections to the destination services that its clients are requesting.

The dual-role pattern occurs when a single IP address operates as both a server and a client. This single IP address architecture is common in circumvention designs. The pattern creates an anomalous behavioral fingerprint. This fingerprint can be observed by state-level middleboxes and used to reveal relay servers even when traffic is fully obfuscated. As illustrated in Figure 1, the *Proxy* entity uniquely demonstrates this dual role: it is shown acting as a server by receiving an Egress Flow (Proxy as Server) and simultaneously acting as a client by initiating an Ingress Flow (Proxy as Client).

In this paper, we aim to answer the following question: **Can the *dual-role* behavioral fingerprint, exhibited by a *single IP address* operating simultaneously as both a server and a client, provide a distinguishing feature that exposes relay servers even when traffic is encrypted or obfuscated?** To address this question, we model a detection method that classifies such behavior based on these observable patterns and evaluate it using a full day of real network traffic, about 17 TB, from the WIDE project [1]. From the dataset, we extract ground-truth relay servers by fingerprinting packet semantics and identifying known relay protocol such as OpenVPN, WireGuard, and SOCKS. We then apply our detection model across the dataset's flows to assess its accuracy against both ground-truth relays and benign TLS servers. The results show that relays consistently exhibit this distinguishing feature, while its occurrence among benign servers is negligible.

This paper is organized as follows. In section 2, we outline the detection model. In section 3, we evaluate the detection against real network traffic. We discuss the merits and implications of the results in section 4, and conclude in section 5.

2 Design

This section outlines the design of our detection method. We first define the censorship model to establish the censor's capabilities. We then detail the three-stage pipeline used to identify the *dual-role* behavioral fingerprint discussed in the Introduction: 1) **Candidate Selection**: heuristic filtering to isolate candidate relays, 2) **Dual-Role Detection**: identifying linked server-role and client-role activities, and 3) **Suspicion Scoring and Classification**: scoring relays based on the destinations they access.

2.1 Censorship Model

We assume a state-level censor with comprehensive monitoring capabilities over all traffic entering, leaving, and transiting the network. The censor can reliably distinguish domestic from foreign IP addresses and identify IP addresses belonging to Autonomous Systems (ASNs) known to host many **Virtual Private Servers** (VPS), hereafter referred to as **VPS-dense ASNs.** At Internet scale, the censor uses the VPS-dense ASNs as an inexpensive heuristic to filter candidate relays, running the behavioral analysis only on this reduced set while minimizing false positives to avoid disrupting benign services. The censor can also extract domains from flow metadata (e.g., DNS queries, TLS SNI, HTTP Host headers) for all observed traffic. Payloads remain encrypted; therefore, detection relies entirely on metadata and observable flow patterns.

2.2 Stage 1: Candidate Selection

We begin with lightweight heuristics to reduce the search space. Candidate relays are foreign servers hosted in the "VPS-dense ASNs" identified in the censorship model. This heuristic is effective because these ASNs are known to be high-priority targets for censors, who apply more complex and resource-intensive rules selectively against them to detect fully encrypted traffic [9]. These servers are filtered to exclude well-known public services, making them suitable targets for further analysis. The resulting set of candidate IP addresses is then passed to the **dual-role detection stage**.

2.3 Stage 2: Dual-Role Detection

For each candidate relay r, we construct a localized graph $G = (V, E)$ from the flows involving r.

- V is the set of nodes, consisting of the candidate relay ($r \in R$), domestic clients ($c \in C$), and the destination services ($d \in D$) they access.
- E is the set of edges, where each edge represents a network flow annotated with a timestamp.

We define a **Dual-Role Instance (DRI)** as a pair of flows that demonstrates the server-role and client-role behavior within **Observation Window (**W**)**, which serves to bound the temporal correlation:

(1) An **Egress Flow (Server-Role):** $e_{out} = (c, r, t_1)$, where $c \in C$.
(2) An **Ingress/Transit Flow (Client-Role):** $e_{in} = (r, d, t_2)$, where $d \in D$.

Let t denote the timestamp of any captured packet. The timestamp t_1 of the initial server-role packet serves as the **trigger** that starts the observation. The window remains open for the duration W, collecting all associated client-role activities at time t_2. The observation period terminates once the current timestamp t exceeds $t_1 + W$.

This stage acts as our primary filter. Any candidate relay r with zero recorded Dual-Role Instances (DRIs) within this window is discarded as benign, as it does not exhibit the complete, observable dual-role behavior.

2.4 Stage 3: Suspicion Scoring and Classification

The previous stage identified all relays r with at least one Dual-Role Instance (DRI). This stage assigns a final **Relay Suspicion Score (RSS)** to quantify the suspicion level of that activity, based on the *destination domains* (d) of those instances.

Let $H(r)$ be the set of all Dual-Role Instances associated with relay r. Each element $h \in H(r)$ contains the destination service $h.d$.

We classify the domains of these destination services into two concrete categories:

- S_{user}: High-suspicion domains (e.g., user-facing services like news, social media). We assign these a high weight, $w_{high} = 0.9$.
- S_{infra}: Low-suspicion domains (e.g., infrastructure like software updates, API backends). We assign these a low weight, $w_{low} = 0.1$.

We define the **Relay Suspicion Score (RSS)** as the *average suspicion level* of all observed instances. It is calculated in a single step by summing the weights of all destinations and normalizing by the total number of instances:

$$RSS(r) = \frac{\sum_{h \in H(r)} W(h.d)}{|H(r)|}$$

where:

- $|H(r)|$ is the total count of Dual-Role Instances for relay r.
- $W(h.d)$ is the weight (w_{high} or w_{low}) assigned to the destination $h.d$.

This RSS score represents the proportion of suspicious activity. A benign server would likely only contact low-weight S_{infra} destinations, resulting in a low RSS (e.g., $RSS(r) \approx 0.1$). A true relay accessing high-value sites will have a high RSS (e.g., $RSS(r) \approx 0.9$). Relays with an $RSS(r)$ exceeding a threshold τ (e.g., $\tau = 0.5$) are classified as circumvention servers.

3 Evaluation

The objective of this evaluation is to validate the core dual-role hypothesis using a 17 TB dataset of WIDE Project traces [1] collected on April 9, 2025. Due to the nature of this backbone traffic, reliable domain metadata, which is required for our full Stage 3 scoring model, is not consistently available. Therefore, this evaluation focuses on a more fundamental question: does the "dual-role" behavior, in which a single IP address exhibits both server-role and client-role activity, hold in practice as a reliable behavioral pattern? To answer this, we conduct a two-part experiment:

1. Ground-Truth Analysis: We first establish a ground-truth (GT) set of relays by identifying known circumvention protocols (OpenVPN, WireGuard, and SOCKS) within the traces. We then apply our dual-role detection heuristic to this GT traffic to determine if known relays exhibit this behavioral pattern and to measure our True Positive (TP) rate.
2. Benign Traffic Analysis: We analyze the general traffic of TLS flows on port 443, which circumvention servers commonly use to blend with benign traffic. We apply the same detection heuristic to this traffic to measure the False Positive (FP) rate, assessing the model's potential for collateral damage.

This approach allows us to validate the feasibility of our detection premise even with limited metadata.

3.1 Flow Extraction Process

To pre-process the dataset, we extract unidirectional flows from the raw packet captures. For each packet, we extract the timestamp, source and destination IP addresses, source and destination ports, and the transport-layer protocol. Client and server roles are inferred using standard port-based heuristics: the server is the endpoint operating on a well-known or registered destination port, whereas the client is the endpoint using an ephemeral source port. Each flow is normalized into a unidirectional 5-tuple *(client_ip, server_ip, client_port, server_port, protocol)*, maintaining this orientation for both communication directions. Packets that do not satisfy these conditions are excluded. For each resulting flow, we enrich both endpoints with metadata from the IPInfo-Lite database [4], including country and organization attributes. This pre-processing stage produces a comprehensive structured flow dataset that represents all observed transport connections and serves as the foundation for our dual-role behavioral analysis.

3.2 Ground Truth Relay IP Addresses

To construct a reliable ground truth for validation, we apply a separate extraction process to the same packet captures. Using *TShark* [8] protocol filters, we identify packets associated with known circumvention protocols, specifically OpenVPN, WireGuard, and SOCKS. For each matched packet, we infer client and server roles using the same port-based heuristic defined in the previous subsection. The set of server IPs identified through this process constitutes our ground-truth dataset, which serves as the benchmark for evaluating the dual-role detection model.

3.3 Dual-Role Detection and Analysis

With the structured flow log (Section 3.1) and the ground-truth set (Section 3.2) established, we now apply the dual-role detection heuristic. Foreign servers are defined as endpoints located outside Japan, as determined by IP geolocation metadata. An *Egress Flow* is a connection initiated from a Japan-based client toward a foreign server, while an *Ingress Flow* is a connection initiated from a foreign client toward a Japan-based server.

All foreign servers observed in the flow log are partitioned into two mutually exclusive traffic types:

- **Ground-Truth (GT) traffic:** foreign server IPs present in the ground-truth set.
- **Benign TLS traffic:** foreign servers that received connections on TLS port 443 but do not appear in the ground-truth set.

The dual-role detection heuristic is applied to both traffic types. A foreign server r is flagged if it satisfies two conditions:

1. It appears as a server in an Egress Flow, and
2. It subsequently appears as a client in an Ingress Flow where the destination port is 80 or 443.

This behavioral condition captures the dual-role pattern in which the same IP address alternates between server and client roles. The resulting classifications enable direct evaluation of detection accuracy:

- **True Positive (TP):** a relay server successfully classified as a relay.
- **False Negative (FN):** a relay server not classified as a relay.
- **False Positive (FP):** a benign TLS server incorrectly classified as a relay.
- **True Negative (TN):** a benign TLS server correctly classified as a benign server.

3.4 Results

With both traffic types defined and the dual-role heuristic applied, we evaluate the detection outcomes using the ground-truth labels. Among the 414 relay servers identified across OpenVPN, WireGuard, and SOCKS protocols, 96 exhibited dual-role behavior and were successfully flagged (**TP** = 96). The remaining 318 servers (**FN** = 318) did not initiate connections toward Japan on port 80 or 443 during the observation window. These unflagged relays are likely

Table 1: Summary of Dual-Role Detection Results

Traffic Type	Metric	Count	Rate (%)
Relays	True Positive (TP)	96	23.2
	False Negative (FN)	318	76.8
	Total servers	**414**	
Benign	False Positive (FP)	179	0.18
	True Negative (TN)	97,472	99.82
	Total servers	**97,651**	
Overall	**Total servers: 98,065**		**Accuracy: 99.5%**

low-activity nodes or configurations operating as split tunnels, where the relay links private LAN networks without forwarding general outbound traffic.

Among the benign TLS traffic, connections initiated by foreign servers toward Japan were exceedingly rare. Out of 97,651 foreign TLS servers analyzed, 214 were flagged as potential relays (**FP** = 214) for exhibiting client-role behavior by initiating ingress connections on ports 80 or 443. We further examined these 214 flagged servers using an external IPInfo Privacy dataset [4] that labels known relays, proxies, VPNs, and Tor nodes. Of these, 35 (16.36%) were independently classified as relays or VPNs in the external dataset, providing additional evidence that the dual-role heuristic captures real circumvention servers blending within benign traffic.

Considering both traffic types, the overall false-positive rate is low relative to the scale of benign TLS traffic—approximately 0.18% of all servers were incorrectly flagged. The true-positive rate (Recall) among known relays is roughly 23% (96 out of 414), reflecting the proportion of active or observable dual-role behavior. When aggregated, the model achieves an accuracy of about 99.5%, dominated by the large volume of correctly ignored benign servers. These results were achieved without the benefit of the selective filtering and scoring mechanisms defined in our design. Specifically, neither the initial VPS-dense ASN filtering (Stage 1) nor the domain-based scoring (Stage 3) could be implemented due to data limitations; their inclusion would likely enhance detection accuracy further.

4 Discussion

In this section, we discuss the practical feasibility of the dual-role detection model, the architectural weaknesses that enable it, and the broader implications of our findings.

The practical feasibility of this detection model rests on established censorship capabilities. State-level censors possess comprehensive filtering of traffic based on blacklists at scale and selectively apply resource-intensive inspection to high-priority targets such as VPS-dense ASNs to detect evasion attempts. Our three-stage pipeline in section 2 aligns with this reality, using lightweight heuristics (Stage 1) to filter candidates before applying the dual-role analysis (Stage 2). The evaluation in section 3 confirms that this approach is both practical and precise, achieving a negligible 0.18% false-positive rate against benign TLS servers. The physical location of monitoring middleboxes is critical: border vantage points provide greater detection coverage than internal placements, since they are more likely to observe outbound connections generated by foreign relay servers.

The fundamental weakness this model exploits is the single-IP architecture common to many circumvention tools, including Shadowsocks, and V2Ray [6]. This design simplifies deployment but introduces an architectural vulnerability. The main defense for these tools is protocol obfuscation that assumes that hiding link semantics is sufficient to prevent detection. Our findings show this assumption is flawed: obfuscation of the link is an insufficient defense when the architecture itself provides detectable signal, where a single IP acts as both a server and a client. This host-level signal cannot be concealed by link obfuscation. While the client-to-relay connection is obfuscated, the relay must generate new, observable flows (via DNS or TLS SNI) to reach end services. A relay accessing user-facing domains (e.g., news or social-media sites) produces a clear behavioral contrast to a benign server accessing infrastructure domains (e.g., update mirrors).

5 Conclusion

In this paper, we investigated whether the single-IP architecture of circumvention design creates an observable feature that persists even under full obfuscation. We model this behavior using a dual-role heuristic detection approach and validate it on large-scale backbone traffic. The observed recall of 23% against our ground-truth set demonstrates that the dual-role behavior serves as a practical and measurable signal for detection. The 0.18% False Positive rate is based on the general TLS traffic observed and is likely a conservative figure, as neither the initial VPS-dense ASN filtering (Stage 1) nor the full domain-based scoring (Stage 3) was implemented in this evaluation. This finding is highly significant because we conclude that architectural design, not link obfuscation, defines the true detection surface. The full potential of this approach, utilizing the complete three-stage pipeline defined in our design, would likely yield substantially higher detection accuracy.

For future work, we will investigate multi-hop proxy architectures and specific optimization techniques, such as split tunneling versus global tunneling, to determine if these configurations produce behaviors detectable at scale.

References

[1] Kenjiro Cho, Koushirou Mitsuya, and Akira Kato. 2000. Traffic data repository at the {WIDE} project. In *2000 USENIX Annual Technical Conference (USENIX ATC 00)*.

[2] Sergey Frolov, Jack Wampler, and Eric Wustrow. 2020. Detecting Probe-resistant Proxies. In *Network and Distributed System Security*. The Internet Society.

[3] Sergey Frolov and Eric Wustrow. 2019. The use of TLS in Censorship Circumvention. In *Network and Distributed System Security*. The Internet Society.

[4] IPinfo, Inc. 2025. IP Data Intelligence. https://ipinfo.io/. Accessed on 2025-11-07.

[5] Aaron Ortwein, Kevin Bock, and Dave Levin. 2023. Towards a Comprehensive Understanding of Russian Transit Censorship. In *Free and Open Communications on the Internet*.

[6] Project V. 2025. Project V: Project V Official. https://www.v2ray.com/. Accessed on 2025-11-07.

[7] Ram Sundara Raman, Prerana Shenoy, Katharina Kohls, and Roya Ensafi. 2020. Censored Planet: An Internet-wide, Longitudinal Censorship Observatory. In *Computer and Communications Security*. ACM.

[8] TShark. 2025. Manual Page. https://www.wireshark.org. Accessed on 2025-11-07.

[9] Mingshi Wu, Jackson Sippe, Danesh Sivakumar, Jack Burg, Peter Anderson, Xiaokang Wang, Kevin Bock, Amir Houmansadr, Dave Levin, and Eric Wustrow. 2023. How the Great Firewall of China Detects and Blocks Fully Encrypted Traffic. In *USENIX Security Symposium*. USENIX.