NO SIEVE IS GOOD ENOUGH: EXAMINING THE CREATION, ENFORCEMENT AND EVASION OF GEOFILTERS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Computer Science

By

Rahel A. Fainchtein, M.S.

Washington, DC
September 5, 2023
No Sieve is Good Enough: Examining the Creation, Enforcement and Evasion of Geofilters

Rahel A. Fainchtein, M.S.

Dissertation Advisor: Micah Sherr, Ph.D.

Abstract

IP-geolocation, or the estimation of a target computer’s geographic location based on its IP address, is widely used across the Internet. Specifically, it is often used (in conjunction with other tools) for the authentication, authorization checking, or fingerprinting of its targets. One particularly noteworthy application of this is geoblocking or geofiltering, in which website owners only allow connections with incoming requests they determine to originate within any of the predetermined geographic regions for which they provide service. Put differently, servers that geoblock (or geofilter) predetermine a set of regions from which their websites will be available. Upon receipt of incoming connection requests, these servers use IP-geolocation to estimate the geographic origin of each one. The servers then permit requests whose estimated geographic origin are within the allowed regions, and deny those found to come from outside of them.

Given the increasing prevalence of geoblocking, many would-be targets of the underlying IP-geolocation seek to circumvent this check, often by spoofing their location, in order to gain access to an otherwise unavailable resource. This has led to a cat-and-mouse dynamic in which the geoblocked targets seek new means to hide or spoof their location, and those performing the check try to crack down on misbehaving targets.

This thesis seeks to better understand this dynamic by performing a multifaceted study of IP-geolocation and geoblocking. In particular, we focus on (1) common mechanisms and channels used for IP-geolocation and geoblocking, (2) geoblock evasion tools that have emerged in response to the increase in geofiltering, and (3) their respective impacts on blocked users’ online security and privacy.
To form a foundation, we study the landscape of IP geolocation techniques, surveying them from the literature and in deployments. We seek to improve our understanding of how these systems currently operate and their limitations. To do this, we review most widely adopted approaches’ (respective) accuracy on both a global scale, and across different geographic regions, and then assess the impact of each mechanism’s accuracy on the utility of its most common applications.

Unlike IP-geolocation, IP-geolocation evasion has not been as comprehensively examined in the open literature. Currently, most studies of IP-geolocation evasion have focused on the circumvention of nation-state censorship. However, few studies have focused on other use cases such as the bypass of server-side geoblocking, or the denial of connections by certain web servers to incoming requests whose IP-geolocation places them outside a predetermined geographic region. Therefore, a main focus of this dissertation is to achieve a more comprehensive understanding of IP-geolocation evasion from perspectives that have not been previously investigated in the literature. Here we seek to understand what these additional motivations are and the methods used to circumvent IP-geolocation across these goals. For each approach studied, we focus on when and where these evasion techniques are effective, if at all, and their respective security and privacy properties.

To help shed light on the IP-geolocation evasion landscape, we perform an in-depth case study of Smart DNS (SDNS) systems. Mainly intended to circumvent geographic access restrictions to online content, SDNS advertises the ability to “unblock” domains that perform geoblocking. This thesis presents one of the first academic studies of SDNS services. We identify a number of serious and pervasive privacy vulnerabilities that expose information about the users of these systems. These include architectural weaknesses that enable content providers to identify which requesting clients use SDNS. Worse, we identify flaws in the design of some SDNS services that allow any arbitrary third party to enumerate these services’ users (by IP address), even if said users are currently offline.
Finally, given these findings, we seek (1) to ascertain users’ motivations for evading geoblocking and (2) to better explain how they perceive and understand the tools available for doing so. We focus on users of one specific tool that is only capable of IP-geolocation evasion, and find that many of these tools’ users do not understand the mechanisms by which their evasion service operates, and how their operation differs from more privacy-preserving technologies. We show that this problem is especially pervasive amongst users whose providers also offer alternate means of IP-geolocation evasion. What’s more, we find that many users have numerous misconceptions about the service’s security and privacy properties. As such, many of its them may place undue trust in the service purveyors, and in the protection their advertised security measures offer.
ACKNOWLEDGMENTS

The work described in this thesis was partially supported by the National Science Foundation under grants DGE-2138078, CNS-1925497, CNS-1718498, and CNS-1149832. The findings and opinions expressed in this thesis are those of the author and do not necessarily reflect the views of the National Science Foundation.
# Table of Contents

## Chapter

1. **Introduction** ..................................................... 1  
   1.1 Background: IP-geolocation ................................. 1  
   1.2 Exploring the Status of Commercial IP-geolocation ...... 3  
   1.3 Understanding IP-geolocation Evasion ..................... 4  
   1.4 Research Questions ........................................... 5  
   1.5 Organization .................................................. 6  

2. **Background and Related Work** ............................... 7  
   2.1 IP-geolocation ................................................. 7  
   2.2 The Role of IP Address Turnover ............................ 11  
   2.3 Geoblocking .................................................... 14  
   2.4 User Perceptions of Geoblocking and its Evasion ......... 15  

3. **Measuring the Current State of IP-geolocation** ........... 19  
   3.1 Industry Uses of IP-geolocation ............................ 20  
   3.2 CDNs, Geoblocking, and Centralization Across the Internet 22  
   3.3 Consistency of Geolocation ................................... 23  
   3.4 Measuring Geofeed Expansion ............................... 25  
   3.5 Comparing Geofeed Results and Commercial IP-geolocation Estimates ... 32  
   3.6 Discussion .................................................... 38  
   3.7 Summary ...................................................... 40  

4. **An in-Depth Exploration of Smart DNS “Unblocking” Systems** 43  
   4.1 Background on DNS .............................................. 45  
   4.2 Background on SDNS’s Proxying Behavior .................. 46  
   4.3 Architecture of SDNS Services .............................. 47  
   4.4 Identifying SDNS Proxies ..................................... 53  
   4.5 System and Threat Models .................................... 55  
   4.6 Privacy Vulnerabilities in SDNS Designs and Implementations 61  
   4.7 Susceptibility to Eavesdropping ............................. 69  
   4.8 Authentication and Authorization Failures .................. 73  
   4.9 Information Leakage through DNS Probing .................. 75  
   4.10 Estimating the Number of Customers and Revenue .......... 80  
   4.11 Discussion ................................................... 82  
   4.12 Ethical Considerations ...................................... 83  
   4.13 Summary ..................................................... 85
List of Figures

3.1 IP-geolocation Analysis Workflow Diagram .......................... 25
3.2 Geofeed IPv4 Space Coverage in Number of IP Addresses ........... 28
3.3 Total Distinct Geofeed ASNs ........................................ 29
3.4 Number of Organizations Estimated as Geofeed URLs ............... 30
3.5 Maxmind Overlap with Geofeed and Overall Accuracy ................ 35
3.6 Maxmind Error Distance CDF Overlay ................................ 36
3.7 Maxmind Inaccuracy Rates Compared to Geofeed Results ............. 37
3.8 IPgeolocation.io Overlap with Geofeeds and Overall Accuracy ....... 38
3.9 IPgeolocation.io Error Distance CDF Overlay .......................... 39
3.10 IPgeolocation.io Inaccuracy Rates Compared to Geofeed Results ... 40
4.1 Workflow of the Operation Phase of an SDNS Service ................. 48
4.2 SDNS IP Enumeration Attack ........................................... 64
4.3 SDNS Cache Enumeration Results .................................... 78
5.1 Top Factors in Participant Choice of SDNS Provider .................. 100
5.2 Participant Perceptions of SDNS Providers’ Overall Trustworthiness (M19) . 102
5.3 Participants’ Trust of SDNS Providers, Broken Down by Actions (M21) . 103
5.4 Participant Understanding of SDNS Functionality ........................ 105
5.5 Perceived Visibility of Internet Browsing to SDNS Provider ........... 106
5.6 SDNS Perceived Security and Privacy Impact ............................ 107
5.7 Participant Beliefs on SDNS Security and Privacy Qualities ........... 108
5.8 Perceived Ethics and Legality of SDNS Usage .............................. 109
LIST OF TABLES

3.1 Top Ten CDNs ................................................. 24
3.2 Geofeed: Top Ten Most and Bottom 21 Least Represented Countries ........ 31
3.3 Geofeed Geographic Representation by Continent ............................... 32
3.4 Maxmind-Geofeed Pull-Date Pairings ........................................... 33
4.1 Identified SDNS Providers .................................................. 56
4.2 Top Ten Countries with the Most SDNS Resolvers ............................ 57
4.3 Top ASes with the Most SDNS Resolvers ..................................... 57
4.4 Top Ten Countries with the Most SDNS Proxies ............................... 57
4.5 Top ASes with the Most SDNS Proxies ....................................... 57
4.6 Summary of Attacks ......................................................... 58
4.7 Mean Number of ASes Traversed per DNS Network Path (by Region) ........ 70
4.8 SDNS Authorization Failures Leading to Open and Universal Proxys .......... 76
4.9 SDNS Most Popular Channels (Derived) ...................................... 87
4.10 Estimated SDNS Resolver Usage and Provider Monthly Profit ............... 88
5.1 Study Participant Demographics .............................................. 96
5.2 SDNS Providers Used by Participants ......................................... 97
This thesis examines IP-geolocation and geoblocking from several different perspectives. Specifically we (1) canvas the existing and deployed IP-geolocation mechanisms and assess their accuracy, (2) examine tools and techniques for bypassing geoblocking, or the filtering of content based on IP-geolocation, and (3) analyze users’ perceptions of these circumvention systems. We evaluate a newly available source of ground truth IP-geolocation data, assess its impact on the existing IP-geolocation and geoblocking landscapes, and show that some geoblock evasion tools can pose serious security and privacy risks to their users. Despite this, we further demonstrate that IP-geolocation evaders that use these services are prone to misunderstanding these tools’ impacts on their security and privacy, and to unduly trusting purveyors of these evasion services.

1.1 BACKGROUND: IP-GELOCATION

At a high level, IP-geolocation is the estimation of a computer’s geographic location based on its IP address. This estimation is generally done (i) using active empirical measurements, (ii) more passively, by querying a database of IP addresses’ registered geolocation information, or (iii) by using a heuristic combination of both passive and active approaches.

Active approaches use delay based measurements to determine a region within which the target IP is most likely located. Among the most commonly used approaches is multilateration, which relies on round trip time (RTT) measurements taken from various anchors, or vantage points distributed across multiple physical regions, and whose (precise) geographic
locations are all known. To estimate the target IP address’s geolocation, the anchors collaboratively complete the following general steps. (1) First, they calculate the current rate of overall network congestion at the time of measurement, and, using that information calibrate a baseline range of speeds at which a message could traverse the Internet. (2) Once the range of possible velocities has been calibrated, a subset of anchors probe the target IP address. Here each anchor measures the RTT between itself and the target. Using the range of message velocities determined during calibration, they then estimate the maximum and potentially the minimum distance the message could have traveled and draw disks or rings denoting the area whose radius is within the range of potential distances traversed. (3) To determine the area in which the target is most likely located, researchers then calculate the intersection of the regions drawn. To achieve higher accuracy, steps (2) and (3) are often run multiple times. At each new iteration, the subset of anchors chosen are those closest to the previous iteration’s result that are still evenly distributed across the geographic sub-region [120].

As described by Muir et al. and Poesse et al., more passive approaches tend to be based off of databases (DBs) that are either publicly available or built by private companies offering IP-geolocation services (such as MaxMind and IP2Location) [93, 99]. As a general rule, the exact mechanisms by which private companies construct their databases tend to be proprietary. However, many researchers have concluded that these databases are often built by scraping or querying, and then aggregating more readily available sources of mappings from IP addresses to geographic locations [99]. These include (but are not limited to):

1. **WHOIS databases**, which are public databases maintained by the different Regional Internet Registrars (RIRs) and contain the registration records determining which entity is responsible for each block of IP addresses allocated as well as their headquarters’ (physical) address and their contact information [66].

2. **DNS LOC records** [33], which, when included within a DNS response, provide additional geographic information about the server running the requested domain. (How-
ever, it is important to note that the LOC piece of a DNS response is optional, and that most DNS authoritative name servers omit them.)

3. **Domain names** which, as Muir et al. note, can contain codes denoting their geographic location.

4. **Routing Data**: In cases where the target’s domain name does not give any hints about its location, Muir et al. point out that its general location can be inferred from the domain names of the closest hops along a packet’s route to the target. As such, one may be able to infer a target’s location with help of the ICMP traceroute function [93].

As described in more detail in Chapter 2, both passive approaches to estimating IP-geolocation and those based on empirical measurements have distinct drawbacks. However, to the degree that they are accurate, these mechanisms provide useful insight into the type of traffic traversing the Internet as well as the status of the Internet’s infrastructure. As such, they have been used for an increasing number of applications. Amongst these are measurement oriented/analytic uses such as measurement of Internet infrastructure and monitoring of end user behavior, as well as more active ones, which tend to respond to IP-geolocation results in real time, and are often geared towards user authorization checks and intrusion detection or prevention.

In this thesis, we focus on geoblocking, an active use of IP-geolocation estimates through which a website owner rejects incoming connection requests it finds to be from a disallowed geographic region, and on Internet users’ attempts to circumvent this type of access restriction.

1.2 **Exploring the Status of Commercial IP-geolocation**

As highlighted in various Content Distribution Networks (CDNs) documentation [14, 28], web servers, or the applications they run generally rely on frequent queries of commercial IP-geolocation services and can process the results in several ways. These range from logging
(high level) traffic source geolocation information in system logs for potential analysis in more passive cases, to the customization of routing, server or application behavior based on IP-geolocation query results in more active ones. Given the high rate upon which commercial IP-geolocation services are relied, knowledge of their current accuracy levels and of the processes through which they construct their underlying databases’ (DBs) mapping between IP addresses and their estimated geographic locations is crucial to understanding the impacts of their results’ applications.

Studying this is especially important in light of the introduction of IP-geolocation feeds, or *geofeeds* in RFC 8805 [77] and their recent standardization in RFC 9092 [103]. As defined in both RFCs, geofeeds provide a framework through which network operators can self-publish coarse-grained geographic information about the IP addresses they control in real-time, and which commercial IP-geolocation services (and third parties) can then query at scale. Put differently, geofeeds provide a new potential source of ground truth geolocation data and could have a profound impact on the state of commercial IP-geolocation. In light of this, we study geofeeds in more depth by (1) examining the extent to which their publication and use has proliferated across the Internet, and (2) assessing their effect on commercial IP-geolocation services and the servers that use them. We provide a more in-depth explanation of geofeeds in Chapter 2 and explore their growth and impact in Chapter 3.

1.3 Understanding IP-geolocation Evasion

Somewhat unsurprisingly, the ubiquity of IP-geolocation’s use, particularly for geoblocking, has also given rise to incentives for would-be targets to evade it. However, there is a limited breadth of study on the means by which this evasion takes place within the open literature. In this thesis, we help broaden understanding of IP-geolocation evasion by focusing on Smart DNS (SDNS), an evasion system with limited prior study. Designed to be lighter weight than other solutions, SDNS relies on its Domain Name System resolvers (SDNS resolvers) improperly resolving customer resolution requests for certain geoblocking domains,
often referred to by providers as supported channels. Instead of returning the correct IP resolution, SDNS resolvers return the IP address of a transparent HTTP proxy they control that is located within the domain’s geofence, or the geographic region(s) from which it allows incoming connections, as determined via IP-geolocation. Within this thesis we explore the technical architecture of these systems as well as their security and privacy properties. We find SDNS systems to have pervasive privacy vulnerabilities due to both faulty system configurations and these systems’ inherent construction.

Looking beyond the evasion systems themselves, we seek to ascertain how users perceive and understand them. Based on both our own and others’ [59, 75] observations of IP-geolocation evasion services’ usage of potentially misleading advertising or verbage across their websites and paid ads, we perform a user study to assess the extent to which users of these evasion systems are mislead, or misunderstand these systems capabilities and limitations and how using them impacts their online security. Given our findings of privacy vulnerabilities within SDNS systems, we once again focus on SDNS users.¹ We find widespread user confusion over the differences between various evasion systems’ respective constructions and functionalities. Given these flaws within users’ perceptions, we observe that many of them put undue trust in the security and privacy benefits these systems purport to offer and do not recognize the threats these systems’ use can pose to their online security and privacy.

1.4 Research Questions

This thesis addresses the following research questions:

1. What is the current state of geoblocking?

As described in existing work by McDonald et al. [17] and as we review in more detail in Chapter 3, many popular websites tend to use Content Distribution Networks (CDNs)

¹This study was originally done in collaboration with researchers at The George Washington University. It was therefore approved by both universities’ IRBs under approval codes STUDY00002643 and NCR202561 for Georgetown University and The George Washington University respectively.
to ensure service reliability, and to rely on their geofiltering functionality to enforce geoblocking. As such, we use this insight to study the IP-geolocation that underlies CDN-based geoblocking to gain additional insight into geoblocking’s efficacy and on its real-world impact on Internet users. To do this we ascertain which CDNs popular websites frequently use, determine the commercial IP-geolocation services upon which these CDNs rely, and measure the IP-geolocation services’ respective rates of agreement with geofeeds over time.

2. How is geoblocking evaded?

We perform a case study of SDNS, a tool capable of geoblock evasion. We study it from both systems and usable security perspectives, to better understand its security and privacy properties.

3. How do users of geoblock evasion tools understand and perceive them?

To better understand the indirect impacts of geoblocking on Internet users, we perform a survey-based user study of \( n = 63 \) users of SDNS services to better understand their perceptions of their SDNS service providers, the SDNS services themselves, and more broadly of geoblocking on the Internet.

1.5 Organization

The rest of this thesis is organized as follows: Chapter 2 describes related work and gives additional background on IP-geolocation; Geofeeds and The landscape of IP-geolocation are covered in Chapter 3; Chapter 4 focuses on Smart DNS, a tool for IP-geolocation evasion with limited previous study; A case study of user perceptions of a particular IP-geolocation evasion tool is presented in Chapter 5; Chapter 6 closes this thesis with a discussion of our findings and conclusions.
We begin by summarizing existing literature covering the different components and facets of geoblocking and geoblock evasion. In so doing, our focus spans across the building blocks of geoblocking, such as IP-geolocation, and industry's implementations and usage of this functionality, geoblocking itself, and contributing human factors, such as users' motivations for adopting tools geared towards geoblock evasion.

2.1 IP-geolocation

Significant work has been done to try and make IP-geolocation more accurate and scalable [6, 27, 37, 55, 111]. However, in its simplest forms IP-geolocation tends to be performed using primarily passive, or database-based approaches such as querying data from WHOIS [66], or other publicly accessible sources, and more active approaches that use empirical measurement methods such as multilateration to determine the bounded geographical region in which the queried IP address must reside.

2.1.1 Active Measurement

Several different flavors of multilateration have been studied in the literature. These approaches, which each make different baseline assumptions about how messages traverse the Internet, have shown varying levels of accuracy when used in different regions. As Weinberg et al. note, their variance in accuracy is related to their general complexity and the extent to which their assumptions align with the state of the regional network being measured. They explain that the relationship between the distance a datagram has traveled, and the
delay until it arrives at a geolocation target, is not straightforward [120]. In particular, they, Candela et al. [37], Arif et al. [83] and others point out, that many factors influence the total delay observed during measurements. These include the circuitousness of a given network path through which datagrams traverse, the number of intermediate nodes or hops along traversal paths, and the bandwidth and latency of intermediate nodes within these paths. Among the most common approaches models take to approximate the delay distance relationship are the assumption of a maximum, and potentially a minimum distance traffic could travel given the observed delay. In simpler approaches, the algorithm graphs the calibration data points as a scatterplot denoting observed delay as a function of the total distance traveled, and finds a best fit of the observed measurements. Using that fit, it then estimates the bounds of how far subsequent probes between the anchors and the target could have traveled [21, 61, 120]. Among the algorithms for which this approach is used are Constraint-Based Geolocation (CBG), which finds a best fit linear approximation of the calibration data [61], and Octant, which finds the convex hull covering 50 – 75% of the calibration data points [21].

On a global scale, Weinberg et al. find that simpler models tend to be more accurate when estimating targets’ geolocation [120]. However, Candela et al. [37], who study how to achieve the highest accuracy with empirical measurement-based geolocation, find that using a single multilateration method on a global scale often translates to higher overall error in the resulting geolocation estimates. They explain that the distance/delay conversion correlates to the average hop length, or the mean distance traveled towards the target in a single hop. Put differently, in regions where connected nodes are further apart, fewer nodes contribute queueing or routing delays over a given distance travelled. This in turn, means that the observed delay is more closely correlated with an increase in distance between a probe and its target. However, in regions with a higher density of nodes, a datagram will likely need to stop at more nodes to cover a given distance, and, as a result, will encounter additional delays from queueing and routing overhead, and potentially from traversing a
more circuitous network path [37]. By using different multilateration approaches based on each region’s observed network topology, Candela et al. find, the resulting IP-geolocation estimates will be within tens of kilometers of their actual locations [37].

Despite their ability to better reflect the state of IP addresses’ allocations at time of measurement, active probing, or empirical measurement based IP-geolocation techniques still pose significant drawbacks. These include their lack of scalability across the IPv4 and IPv6 namespaces, the potential for interference due to high levels of cross traffic and/or network congestion, as well as regional variation in rates of targets’ responsiveness to ICMP and traceroute messages, which are most commonly used for probing [20, 98].

2.1.2 Passive IP-geolocation

Given the high overhead cost in active, or probing based methods of IP-geolocation, both researchers and industry have also turned to more passive mechanisms of measuring IP address usage and geolocating IP addresses across the Internet. While we note that the exact geolocation methodologies used by commercial IP-geolocation services are proprietary, previous work has inferred that they likely are based off of publicly available resources such as WHOIS, which have been found to have high rates of errors [93, 99]. In contrast however, researchers have found that passive analysis of network traces and system logs collected at privileged vantage points in the Internet, provide significant insight into the state of IP address utilization as well as the overall address churn that takes place [32, 98]. In particular, Richter et al., who examine a full year’s worth of system logs from a major commercial CDN, note their ability to observe address turnover patterns at the granularity of individual IPv4 addresses [98]. Dainotti et al. also find that passive measurements collected from a privileged vantage point, (in this case network traces and bidirectional network flows,) gives significant insight into the state of address utilization [32]. Despite these findings, passive approaches to IP-geolocation still present several disadvantages. For commercial IP-geolocation databases, these commonly include record staleness, lack of data
granularity, and lack of data authentication [93]. Datasets originating in system logs, network flow logs, and/or large-scale network traces must be collected from Internet vantage points that can observe high volumes of traffic, and often contain numerous spoofed addresses, which must then be identified and filtered out [32, 98]. Moreover, if not collected recently, these traces may contain stale records. Despite these potential drawbacks, Gharaibeh et al. [84], who measure commercial IP-geolocation services’ ability to accurately locate Internet routers, find high rates of country level interservice agreement across the six commercial IP-geolocation services analyzed. However, they note that the commercial datasets’ show between 77.5% - 89.4% country-level accuracy over the routers included within their respective databases – a significantly lower country-level accuracy than these services respectively advertize [84]. Gharaibeh et al. further caution that these services’ country-level accuracy varies widely across different countries and regions\(^1\). For these reasons Gharaibeh et al. recommend against the use of commercial IP-geolocation services for the identification of routers or Internet infrastructure whenever possible [84].

2.1.3 Hybrid IP-geolocation Approaches

Hybrid IP-geolocation approaches primarily rely on one of the two approaches listed above as an initial starting estimate of the target’s location. However, they then refine them by augmenting outside information which is often collected using the other type of approach. This refinement is often achieved by joining and/or aggregating the information gleaned from both types of measurements to build a better informed set of estimate constraints. Given database-based approaches’ limited ability to accurately locate server infrastructure, hybrid approaches, particularly those primarily based off of passive geolocation techniques, are frequently designed and used for that expressed purpose. Some examples of this are Client Centered Geolocation (CCG) [27] and RIPE IPmap [6]. For CCG, user endpoint geolocation

\(^1\)e.g. While most commercial providers showed over 90% accuracy in identifying routers in the USA, most providers showed between 20% and 39% accuracy when locating routers in Canada [84].
estimates are initially obtained by querying IP-prefixes in the MaxMind Geolite database. These estimates are then refined using the following additional heuristic constraints:

1. The underlying Geolocation DB must have city-level results for the prefix;
2. A prefix may not have servers in multiple locations; and
3. Prefix location estimates given by the database(s) may not fall outside of the feasible region identified by the initial ping measurements and the application of the speed of light constraints [27].

In comparison to CCG, RIPE IPmap estimates geolocation from a broader set of data inputs. Specifically, it takes multiple sets of active measurements collected via multilateration, reverse DNS query results, crowdsourced geo-tagged data, and queries to the PeeringDB system [6]. Responses are given a numerical score which respectively estimates the system’s confidence in their correctness. Unless otherwise specified within the query, IPmap returns the result with the highest score [6].

While RIPE IPmap’s rate of accuracy is still an area of active research, preliminary evaluations seem to indicate it having a high degree of accuracy. Iordanou et al. find that, among the subset of IP addresses that overlapped with their own ground truth measurements, roughly 99% of them matched [30].

### 2.2 The Role of IP Address Turnover

Regardless of the means by which it was collected, all static IP-geolocation data is prone to becoming stale after a certain period of time. However, the rate at which geolocation records and IP address reputations become stale is largely dependent on how and when DHCP churn, or the reassignment of IP addresses occurs. As found by many previous studies, this can be difficult to characterize as DHCP churn varies widely across different Internet Service Providers (ISPs) and geographic regions [26, 96, 98].
Richter et al. find that client networks contribute to this churn to varying degrees. In particular, the degree of a client network’s impact is dependent on the size of its IP address pool, and on the ISP’s IP address allocation and reassignment practices [98]. In reference to the latter, Padmanabhan et al. note this behavior is largely determined by the protocol the ISP uses to dynamically allocate the IP addresses they control. They explain that DHCP [36, 80] is generally configured to try and keep IP address allocations stable, whereas the use of PPP [89, 97] is set to ensure more efficient use of limited IP addresses available, and, as such, will often reassign the IP address after a session is completed [96].

As they, and others observe, these tend to vary regionally. Specifically, they find IP address allocations in North America to be the most stable [26, 32, 96, 98]. In fact, Padmanabhan et al. find that customers there often maintain the same IP address for weeks [96]. In contrast, researchers observe that ISPs in numerous other countries and regions enforce periodic address renumbering after a customer has used an IP address for some predetermined session length, often a multiple of 24 hours [96].

However, despite the prevalence of periodic renumbering (particularly in regions to which fewer IPv4 addresses are allocated), the impact of these frequent address reassignments on (coarse grained) IP-geolocation, or general online tracking, is likely limited. Richter et al. observe, these renumberings often cover addresses spanning /31 or /32 prefix masks, meaning that a significant number of IPv4 addresses remain within the same subnets despite ISPs’ widespread use of periodic address renumbering [98]. In fact, Mishra et al. [90], who perform a user study in which they track individuals’ public IP addresses over time, find participants frequently reuse certain public IP addresses over time, find addresses within a specific /24 block, for periods lasting longer than 30 days. Moreover, they find these IP addresses (or prefix masks) are so consistent that they can be used to fingerprint and later re-identify these individuals as they browse the Internet [90].

As both Richter et al. and Padmanabhan et al. point out, to have a greater impact on applications like geolocation, the IP address churn must span larger address spaces (or
shorter prefix masks). This more widespread churn, they explain, is generally associated with network restructuring events, and can be related to major disruptive events such as network or power outages, or could be the result of pre-planned restructuring within an ISP [96, 98].

2.2.1 Geofeeds

Given the high demand for accurate geolocation information amongst content providers, (and the negative impact stale or inaccurate information can have on legitimate users’ access to their content,) there is a need for these services to obtain up-to-date geolocation information from ISPs. However, anecdotal evidence suggests content providers’ IP-geolocation estimates often miscalculate the geolocation of certain IP addresses, or incorrectly flag them for being VPNs or unblockers [10]. Originally defined in RFC 8805, IP-geolocation feeds (geofeeds) were designed to address this issue [77]. As defined in RFC 8805, these comma separated value (CSV) files were to be self-published by network operators (such as Autonomous systems and ISPs) and to contain announcements of IP address allocation changes [77]. This would allow network operators to announce IP address allocation changes or reassignments directly to content providers rather than wait until they observed them [77]. The content providers, in turn would have the opportunity to update their records accordingly, and in doing so, to prevent them from becoming stale, and improve the quality of their online services and websites’ location-based customizations [103].

To aid in locating these feeds and in authenticating their accuracy, RFC 9092 proposes a standard which outlines how each Regional Internet Registry (RIR) should store the list of domains at which all of its network operators’ (respective) geofeeds were published [103]. Additionally, it further describes how to locate the geofeed file that contains a given IP address, how to authenticate geofeeds, and how they can be fetched from each RIR [103].
2.2.2 What Does Industry Do?

While RFC 9092 notes that “geolocation providers have bulk WHOIS data access at all the RIRs” [103], the extent to which many content and service providers and IP-geolocation services currently rely on geofeeds is not publicly known. However, there is significant evidence indicating that industry frequently relies on commercial IP-geolocation services to estimate the geographic origins of incoming requests. This is particularly prevalent amongst CDNs, which often note their reliance on these within their documentation, or home pages\(^2\), and which McDonald et al. [17] find to be widely used by popular websites.

2.3 Geoblocking

Despite seemingly widespread use of geoblocking, the behavior has not been very thoroughly studied. Tschantz et al. [116], who perform a large scale measurement of the practice, find that it appears to be ubiquitous. In particular, they note widespread blocking of IP addresses associated with developing countries by websites hosted in industrially advanced ones, such as the United States and other European nations [116]. McDonald et al. track CDN-based geoblocking by sending requests to different CDN-supported websites from hundreds of vantage points around the world [17]. They find that geoblocking occurs across a wide range of countries and websites, and is implemented for a wide range of reasons. These include (but are not limited to) compliance with legal or diplomatic restrictions such as economic sanctions, export control legislation, and copyright usage restrictions [17]. To this end Ramesh et al., who measure the network level changes that occurred in the weeks following Russia’s 2022 invasion of Ukraine, find that 45% of the Russian government domains tested blocked connection requests that originated outside of Russia and Kazakhstan. They also identify 444 non-Russian websites which began geoblocking connection requests from Internet users based in Russia during that period [101]. Kumar et al. observe that, in

\(^2\)In particular, Cloudflare notes using Maxmind for its geolocation backend [28], both Akamai and Amazon CloudFront use Geoguard [1], and Fastly lists Digital Element as the purveyor of its IP-geolocation utilities [4, 12].
addition to its widespread use in websites and online services, geoblocking is also common within the Google Play store, and by extension within the Android app ecosystem. Similar to its counterpart for websites, they find multiple forms of geoblocking within the ecosystem and note that various entities can stipulate geo-restricted access or request that it be implemented. Amongst the parties they note can stipulate geographic restrictions to application access, or request that such limits be implemented are application developers and government entities. Contrary to their own hypothesis, they find that the majority of observed geo-differences and georestrictions are requested by application developers rather than by takedowns requested by specific governments or due to apps’ breach of local laws [79].

2.4 User Perceptions of Geoblocking and its Evasion

Unfortunately, the lack of research about geoblocking also encompasses user perceptions of the practice. In particular, questions such as how Internet users cope with geoblock-induced limits to their Internet browsing, and how these in turn impact their online security and privacy, have not been extensively studied. However, existing research has shed light on factors that likely contribute to Internet users’ perceptions of geoblocking [17], and, for those who decide to evade it, their perceptions of the tools and services that are able to circumvent it [16, 50, 51, 75, 94, 102, 106, 109, 112]. In particular, McDonald et al. observe that websites that geofilter do not usually announce or publicize that they do so [17], and both Khan et al. and Ramesh, Vyas and Ensafi find that many users are vulnerable to misunderstanding the limitations on the security and privacy guarantees that VPNs, which can evade geofiltering, actually offer [75, 102]. Other works, which focus on Internet users’ assessment of the security and privacy of the offerings they encounter online, or more specifically, on the online offerings that boast geoblock evasion capabilities, help answer some of these questions.
2.4.1 Assessing Online Security and Privacy more Broadly

Several papers note that users do not assess the risks posed by system usage completely logically, and that their decisions to share or withhold sensitive information tend to be contextually based. Acquisti and Grossklags, who study users' attitudes and decision making processes concerning their privacy, explain that users' notions of privacy are complex and multifaceted, and that Internet users face major limitations on their ability to logically assess the implications of their decisions to share or withhold sensitive information [16]. These include users' lack of access to contextual information about the full scope of their decisions’ impacts, as well as their bounded rationality, or their limited ability to synthesize, recall and logically apply all of the available information when making a decision. To cope with these shortcomings, Acquisti and Grossklags note, users rely on a set of mental shortcuts or cognitive heuristics to try and qualitatively approximate the missing information and logic steps, and arrive at a decision more quickly. However, in addition to logical approximations (and often, the simplifications) of the respective scenarios users face, users' decisions are also influenced by their biases and other “psychological deviations from rationality” [16].

Gambino et al. [51] identify several cognitive heuristics that users often resort to when determining whether (or not) they are willing to share private information with a given website. Specifically, they note several positive heuristics that make users feel more at ease sharing more sensitive information, as well as negative heuristics that make users more wary of sharing it [51]. Shyam notes that users tend to assess the trustworthiness of a site based on its appearance and visual presentation moreso than on the site’s content [112].

Despite this apparent paradox, previous work has found that online tracking, another end for which IP-geolocation is used, is widely perceived by users as an invasion of their online privacy. Coopamootoo et al. [78], Kacsmar et al. [19], and Ur et al. [23] observe, many Internet users find third party online tracking to be deeply intrusive, and are concerned about companies obtaining sensitive types of their information such as financial or health data. In more detail, Kacsmar et al., who study users' comfort with different data tracking
policies and behaviors, find that different factors impact the extent to which participants find online tracking and data sharing to be acceptable. Specifically, these factors encompass (1) how users’ consent to tracking or data sharing is initially given, (2) how consent transfers in the events of company collaboration, mergers, or acquisitions, and (3) whether it includes sensitive categories of information, such as health data. Overall, they find participants believed consent to online tracking and to data sharing had to be explicitly given, and that respondents raised objections to being coerced into data sharing to receive services and to the collection of data to increase companies’ revenue [19]. However, Ur et al. [23] find that existing means of notifying users they are being tracked and allowing them to choose when to allow tracking, if at all, are not effective [23]. When it comes to how users’ emotional reactions to unwanted online tracking translate into their taking actions to protect their online privacy, Coopamootoo et al. [78] find significant cultural and gender differences amongst respondents’ perceptions of online tracking, and the protective actions they took as a result. Specifically, they note that women tend to have more intense negative emotional reactions to being tracked, but that men were more likely to take actions to protect their online privacy, such as adoption of security tools designed for that purpose (including, but not limited to ad blockers, Tor [34] and VPNs) [78].

As Khan et al. [75] point out, and our findings confirm, many IP-geolocation evasion tools fail to adequately protect users from this type of tracking [75]. However, users’ online privacy concerns are not unwarranted. First pointed out by Olejnik et al. [82], and later confirmed by Mishra et al. [90] and Bird et al. [108], Internet users can be uniquely identified and tracked using their web browsing history. Worse still, both Olejnik et al. and Bird et al. show that online behavioral tracking is likely very common [82, 108]. In fact, Iordanou et al. [30], who study cross border web tracking in the EU, find that despite regulations such as the General Data Privacy Regulation (GDPR) [31] and the California Consumer Privacy Act [29], cross border tracking covers numerous sensitive, and thus protected data categories [30].
2.4.2 User Perceptions of Geoblock Evasion Tool Security

While not all tools capable of geoblock circumvention help shield users from online tracking, previous work on their perceptions of these tools has highlighted widespread misconceptions about the security they each provide. Gallagher et al. [50] focus on perceptions and usage of Tor amongst both technical experts and non-experts, and find significant differences between them. Specifically, they note, non-experts often think of Tor as a centralized service, conceptualize it more abstractly, and fail to recognize how the actions they take while browsing can increase their risk of deanonymization. In contrast however, Gallagher et al. find that technical experts had much more accurate and detailed understandings of Tor’s decentralized architecture and the deanonymization risks associated with how they browsed the Internet [50].

Additionally, while not the most cited motivation for use of tools capable of IP-geolocation evasion, avoiding online surveillance has been found to be a common motivation for adopting these tools. In particular, respondents in various studies indicate using Tor [50] and VPNs [78, 94] for this purpose. In our research we also find that some users adopt SDNS for this reason despite the tool’s lack of protection against tracking (see Chapter 5). However, as Namara et al. [94] find, most VPN users do not use these services for additional online security and privacy. Instead, they find participants often seek out access to geo-restricted websites. They further note that for participants who chose to use VPNs for a specific task or to fulfill a specific need for anonymity, many decided to discontinue VPN use due to the tools’ lack of usability [94]. In Chapter 5, we report similar findings amongst participants in our user study who reported having previously used VPNs and then having switched to using SDNS. However, our study encompasses the perceptions of IP-geolocation evasion tool users rather than solely focusing on VPN users who turned to these tools to protect their online privacy and security.

In the following chapters, we more closely analyze the SDNS architecture and ecosystem (Chapter 4) and explore users’ perceptions of these services (Chapter 5).
Measuring the Current State of IP-geolocation

The issue of determining users’ or servers’ geographic locations has become a source of increased interest to various entities. Amongst these are services facilitating online financial transactions, copyright holders or licensees, online advertisers and researchers. As such, various means of ascertaining or inferring location have become increasingly prevalent across the Internet. In particular, IP-geolocation, or the estimation of geographic location based on a machine’s IP address has become one of the most common approaches to inferring this information.

Despite the lack of an inherent mapping between an IP address and geographic location, using IP-geolocation has the distinct advantage of being a technique from which an online user cannot easily opt out. Unlike other sources of users’ location information such as HTML headers or a mobile phone’s GPS coordinates, the IP protocol does not allow a user to make a request without sharing their source address\(^1\). Put differently, all IP datagrams must contain valid source and destination IP addresses to allow the two entities to communicate. As such, IP addresses are the only freely available source from which a machine’s geographical location can be inferred. For this reason IP based geolocation (IP-geolocation) is viewed by many as sufficiently reliable.

\(^1\)A user could always use a means of obscuring their IP address such as Tor, a VPN, or a proxy as we will discuss later.
As Marketa Trimble [113] notes, this view is particularly common amongst lawmakers and litigators. So much so that in some cases it has resulted in their mandating that website owners use IP-geolocation (or more specifically geoblocking) to comply with various laws and regulations including those concerning intellectual property, regional gambling restrictions, and oversight of online financial transactions [113].

However, as noted in Chapter 2, lawmakers/litigators’ perceptions of commercial IP-geolocation significantly conflict with research in the literature, which has found that it faces numerous shortcomings. In particular, commercial IP-geolocation services have been found to be largely unreliable when it comes to geolocating Internet infrastructure (e.g. servers and/or routers) [30, 120] and to underrepresent, and more frequently mislocate Internet vantage points in less industrialized countries [99].

Although not the only sources of these discrepancies, the dynamic nature of both Internet address allocation and international relations are common causes for changes in how and where IP-geolocation and geoblocking are implemented. In particular, it is worth noting that geoblocking is often used for compliance with sanctions and other international laws. As such, the degree to which individuals at different vantage points face geoblocking likely changes with current events. Despite this, many aspects of both IP-geolocation, and its applications (particularly geoblocking) still have yet to be studied.

In this chapter, we seek to better understand how the introduction of geofeeds [77] and of their more recent integration with WHOIS database(s) [103] has impacted (and will continue to impact) the general and commercial IP-geolocation ecosystems. Using this information, we then assess how our findings impact geoblocking across the Internet.

3.1 Industry Uses of IP-Geolocation

By and large, how users experience the Internet, has become increasingly dependent on the vantage point from which they access it. Specifically, a growing proportion of this variance in server behavior is determined by the geographic region in which their clients are based,
as estimated by IP-geolocation. This customization has taken on a wide variety of forms including but not limited to:

- Customization of page settings such as its default webpage language preferences;
- Customization to regional markets, which could take the form of regional price discrimination, and provincial variance in product offerings;
- Selective compliance with regional laws: as noted by various network administration blogs, it is not uncommon for webservers to only display statements or warnings outlined in laws such as GDPR, or CCPA, to individuals located within the geographic regions in which they are required [11, 81]; and
- Customization of page access, which is more commonly referred to as geoblocking.

Despite the variance across different types of IP-geolocation-based customization, the quality of these customizations is frequently diminished due to inaccuracies in the underlying IP-geolocation providers’ estimates.

For Internet users who are mis-located, this often translates to a degradation in websites’ services and their overall usability.

As noted in RFC 8805 [77], this breakdown in website service was a large motivation for the definition of self-published geolocation feeds, or geofeeds. They describe the problem as follows:

When an ISP, for example, changes the location where an IP prefix is deployed, services that make use of geolocation information may begin to suffer degraded performance. This can lead to customer complaints, possibly to the ISP directly. Dissemination of correct geolocation data is complicated by the lack of any centralized means to coordinate and communicate geolocation information to all interested consumers of the data [77].

RFC 8805 therefore introduces self published geolocation information feeds, or geofeeds, as a means by which network operators such as Autonomous Systems (ASes) or ISPs can share the actual geolocation data of the IP addresses and prefixes they control.
RFC 8805 defines a geofeed file to be a comma separated value (CSV) file in which each entry contains either a single IP address or range of addresses in CIDR notation and its corresponding city and country-level geographic information.

RFC 9092 [103] expands on this by centralizing the location to which network operators can publish the URLs of their geofeeds. This enables geofeed consumers, such as commercial IP-geolocation providers, to easily find them. To do this RFC 9092 defines a mechanism through which ASes can register their geofeed URL with their respective Regional Internet Registries (RIRs), National Internet Registries (NIRs) or Local Internet Registries (LIRs) and in so doing, add them to their WHOIS database entries. As part of this update, RFC 9092 proposes the expansion of the Routing Policy Specification Language (RPSL), which Internet Registries use to specify registrant information, to include a new geofeed field which holds the URL of the registrant’s geofeed [103].

3.2 CDNs, Geoblocking, and Centralization Across the Internet

While their focus is primarily on geoblocking, McDonald et al. [17] note that this increase in geolocation-based customization is an effect of the increased centralization of the Internet. Specifically, they highlight that many of the most visited websites rely on CDNs for both the increase in their websites' responsiveness and robustness against distributed denial of service attacks, and for other functionalities such as the CDNs' IP-geolocation-based offerings. This in effect, they note, causes more distinct websites to be hosted at least in part on shared servers’ infrastructure. As a result, this often translates to CDNs offering more functionality to all of their customers including those whose subscription is for lower tiers of service. Since this centralization could make geoblocking easily available to website owners who do not have good reason to implement this functionality, or who may be inclined to overblock, McDonald et al. find that it likely poses a threat to worldwide Internet accessibility [17]. In their analysis of Cloudflare data, McDonald et al. specifically focus on when, and against whom Cloudflare customers implemented its geoblocking capabilities. These capabilities, they explain, were
offered through Cloudflare’s Firewall Access Rules feature and allowed customers to allow-
list, challenge (e.g. via CAPTCHA) or block their respective websites’ visitors based on
their Country, IP-address or AS number (ASN) [17]. The dataset, which they received from
Cloudflare, reflected all active country-scoped access rules set by Cloudflare customers in
July 2018. As McDonald et al. explain, this was particularly noteworthy, since it fell during
a five month regression period (April 2018 - August 2018) in which country-level blocking
capabilities were not limited to Enterprise level customers, but instead were temporarily
available to customers of all service tiers [17]. They then observe that many Business, Pro
and Free tier customers had opted to use the feature despite its only having been available for
three to four months at the time of the dataset’s collection. This, McDonald et al. conclude,
implies that when given the option to enable geoblocking, many websites will choose to do
so [17].

3.3 Consistency of Geolocation

Given McDonald et al.’s observations, our study of IP-geolocation was largely motivated
to better understand the mechanisms through which CDNs, or their underlying commercial
IP-geolocation service providers estimate Internet user’s vantage points. As such, we present
a cursory breakdown of the top ten CDNs most used by websites within the Alexa Top
Million as determined by Experte.com [8], and the IP-geolocation providers that support
them in Table 3.1.

In addition to CDNs’ role in centralizing the machines through which Internet traffic
travels, widespread reliance on CDNs also consolidates the set of resources (or in this case
the IP-geolocation providers) that websites consult to estimate the geographic origins of
their incoming requests.

Given the intended role of geofeeds, we hypothesized that the widespread adoption of
their publication by network operators and reference by commercial IP-geolocation providers
would significantly increase interservice agreement on IP addresses’ geographic vantage
Table 3.1: Top Ten CDNs. The top ten most used CDNs as measured by Experte.com [8], and the underlying IP-geolocation provider used to support them, as determined by evaluating their documentation. Citations denote the CDN documentation page(s) or IP-geolocation provider websites from which this information was gleaned.

<table>
<thead>
<tr>
<th>Rank</th>
<th>CDN</th>
<th>Geolocation provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cloudflare</td>
<td>Maxmind [28]</td>
</tr>
<tr>
<td>2</td>
<td>jsDelivr</td>
<td>IPgeolocation.io [67]²</td>
</tr>
<tr>
<td>2</td>
<td>Amazon Cloudfront</td>
<td>Geoguard [1]</td>
</tr>
<tr>
<td>3</td>
<td>Sucuri</td>
<td>unknown</td>
</tr>
<tr>
<td>4</td>
<td>Imperva (Previously Incapsula)</td>
<td>internal solution [9]³</td>
</tr>
<tr>
<td>5</td>
<td>Google CloudCDN</td>
<td>internal solution</td>
</tr>
<tr>
<td>6</td>
<td>Akamai</td>
<td>Geoguard [1]</td>
</tr>
<tr>
<td>7</td>
<td>Netlify</td>
<td>unknown</td>
</tr>
<tr>
<td>8</td>
<td>Fastly</td>
<td>Digital Element [4, 12]</td>
</tr>
<tr>
<td>9</td>
<td>Azure CDN</td>
<td>internal solution [13]</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

points, and as a result, further amplify the effective Internet centralization highlighted by McDonald et al. To test our hypothesis and to better understand these IP-geolocation ecosystems, we set out to address the following research questions:

(1) Since the recent introduction of a centralized geofeed registration system within the WHOIS database, to what extent have network operators registered geofeeds and used them to publish ground truth geolocation data? (2) How, and to what extent has their adoption impacted commercial IP-geolocation services’ overall accuracy? (3) Based on these observations, can we conclude commercial IP-geolocation services actually rely on geofeeds as a source of ground-truth geolocation data, and if so, to what extent?

We describe our work in the next sections as follows:

• Section 3.4 describes our analysis of geofeed adoption;

• We compare geofeed records with overlapping commercial DB entries in § 3.5; and

• Our follow-up analysis is covered in § 3.6.
Figure 3.1: IP-geolocation Analysis Workflow Diagram. A workflow diagram of the analysis processes used to collect and analyze the geofeed records as described in § 3.4 (left), to collect and analyze the commercial IP-geolocation services’ databases as outlined in § 3.5.1 (right) and to compare overlapping results between the two as described in § 3.5.1 (bottom).

3.4 Measuring Geofeed Expansion

To assess the impact of geofeeds on the commercial IP-geolocation ecosystem, we first seek to understand how their coverage expands over time. As indicated in § 3.1, network operators who want their geofeed URLs to be published within their WHOIS (inetnum:) entries can register them with their respective Internet Registries. Put differently, the expansion of geofeed coverage requires network operators to actively opt-into using geofeeds, and to accurately update them when they reallocate addresses to new geographic regions. Therefore we also investigate how and where networks operators have opted in and the extent to which different geographic locations are represented within the geofeeds. In measuring geofeed adoption we sought to answer the following more specific research questions:
1. At what rate are network operators registering and publishing new geofeeds?
2. What is the coverage of geofeed records across the Internet?
3. How has this coverage changed over time?
4. Where are the IP ranges listed within these geofeeds geographically located?
   Are they evenly distributed, or concentrated within a few geographic regions?

3.4.1 Geofeed Measurement Methodology

To answer these questions we use the geofeed finder [86], an open source tool, developed by one of RFC 9092’s authors (M. Candela) that queries current WHOIS records to locate geofeed URLs, pulls the geofeeds contents and verifies their integrity in accordance with the requirements set by RFC 8805 [77] and RFC 9092 [103]. Using the tool we initially queried the geofeed records in April 2022, and, starting in September of that year, pulled updated geofeed records every 13-16 days\textsuperscript{4} over the course of a nine month period (September 2022 - June 2023).\textsuperscript{5} Finally, we used the geofeed results and the tool’s other output files, to measure geofeeds’ coverage of the IPv4 namespace and to evaluate their expansion across different regions. A workflow diagram outlining our methodology is given in Figure 3.1.

3.4.2 Geofeed Measurement Limitations

While geofeed results inform us about their uptake, we acknowledge our study faces limitations due to having to upgrade the geofeed finder during measurement, and due to our inability to fully verify geofeed results’ veracity.

Geofeed-finder Version Changes The software version used between April 2022 thru January 2023 was deprecated during the measurement period and we therefore had to upgrade to release version 1.7.1 on February 1, 2023. In June 2023 we once again had to upgrade

\textsuperscript{4}Measurements were initially pulled manually once every two weeks and were later automated to run on the 13th and 28th of each month.

\textsuperscript{5}While this thesis covers data for a nine month period, we note that as of September 2023 our collection of geofeed data is still ongoing.
to geofeed-finder release version 1.10.0 due to the deprecation of version 1.7.1. While the newer software releases were more effective in capturing and efficiently verifying geofeed results, we were unable to rerun prior geofeed pulls using the newer software. As a result, there may be some inconsistency in the rates of geofeed opt-in between results collected before and after the software upgrades.

**Assumption of Geofeed Reliability**  In our measurements of geofeeds we assume the geofeed entries we obtain from the geofeed finder [86] accurately depict the geographic regions to which its publishers (or rather the network operators) allocate the IP addresses under their control. However, as highlighted by RFC 9092 §7 the only known solution for the full verification of geofeeds’ veracity would be to require that all network operators registering and publishing geofeed files cryptographically signed their geofeeds with RPKI keys. As RFC 9092 explains, in the absence of this requirement, network operators could exploit the weak or missing authentication of numerous RPSL repositories to spoof `inetnum:` entries and set them to point to malicious geofeed files [103]. To illustrate this risk, they describe a specific attack through which this vulnerability could be exploited to effectively dilute the authentication guarantees provided by an RPSL entry with a cryptographically signed geofeed.

If an `inetnum:` for a wide address range (e.g. a /16) points to an RPKI-signed geofeed file, a customer, or attacker could publish an unsigned equal or narrower (e.g. a /24) `inetnum:` in a WHOIS registry that has weak authorization, abusing the rule that the most-specific `inetnum:` object with a geofeed reference **MUST** be used. [103]

While we note that we did not have any means to identify cases in which spoofing occurred, we believe our measurements still provide valuable insights into the commercial IP-geolocation ecosystem.
Figure 3.2: Geofeed IPv4 Space Coverage in Number of IP Addresses. The total IPv4 addresses tripled from 3,220,257 IPv4 addresses (0.08% of the IPv4 namespace), on April 2, 2022 to 9,775,496 addresses (0.23% of the IPv4 namespace) on June 16, 2023.

3.4.3 Results

Growth of Geofeed Adoption Overall, we find network operators’ adoption of geofeeds slowly but steadily grew over the course of the observation period. As shown in Figure 3.2, geofeeds’ IPv4 space coverage tripled from 3,220,257 on April 2, 2022 to about 9,775,496 IPv4 addresses on June 16, 2023. Figures 3.3 and 3.4 show similar patterns across the unique ASNs identified in each geofeed pull and the upper bound on the total unique organizations with registered geofeeds.

---

6In light of RFC 9092 requirement that every cryptographically signed geofeed only be cited within a single WHOIS (inetnum:) entry [103], this heuristic can only provide an upper bound for the total unique contributing organizations within a given geofeed pull.
While all three measurements show significant growth in geofeed opt-in, geofeeds’ coverage over the Internet namespace remains minimal. In particular, as of June 16, 2023, the 9,775,496 IPv4 addresses announced within the geofeeds only account for roughly 0.23% of the IPv4 address namespace. Similarly, the number of unique ASNs identified nearly tripled from 375 ASNs in December 2022 to 1,048 ASNs in June 2023, but this only equates to 0.92% of the ASNs allocated \cite{95}. Given the limited current uptake/opt-in to geofeeds, the extent to which they could serve to help improve commercial IP-geolocation providers’ overall accuracy also appears to be limited.

Despite geofeeds’ overall limited opt-in to date, we observe that network operator opt-in for geofeeds varies widely by geographic region. Table 3.2, which provides a breakdown

\footnote{While we found the estimated upper bound for the total unique organizations contributing geofeeds nearly quadrupled, we were not able to find an authoritative source with the total distinct organizations with IPs in the IPv4 namespace.}
Figure 3.4: Number of Organizations Estimated as Geofeed URLs. An upper bound for the total distinct organizations whose geofeed entries were included in each measurement. This was estimated by counting the unique geofeed URLs whose records were included in the geofeed-finder’s output.

A comparison of the total IPv4 addresses associated with the 10 most and 21 least represented sovereign countries within the geofeeds, shows a strong concentration of the geofeed IPv4 addresses geolocated to wealthier and more industrialized countries that have more Internet infrastructure available. This is particularly well exemplified by the breakdown of geofeed IPs by location in April 2022 - where IPv4 addresses geolocated to the United States account for about 73.7% of all geofeed IPv4 addresses. While by June 2023 numerous additional countries held more substantial proportions of the total geofeed IPv4 addresses, the vast majority (about 8,841,494, or 90.8%) of geofeed IPv4 addresses were located in Europe and North America combined and 4,988,562, (45.6%) and 3,962,930, (39.6%) in Europe and North America respectively. In stark contrast, only 29,330 IPv4 addresses, or 0.30%
Table 3.2: Geofeed: Top Ten Most and Bottom 21 Least Represented Countries.
Top Ten most represented countries within the geofeeds (top) and 21 least represented countries in the geofeed results (bottom).

<table>
<thead>
<tr>
<th>Country</th>
<th>Continent</th>
<th>IPs(4/22)</th>
<th>% Gfeed(4/22)</th>
<th>IPs(6/23)</th>
<th>% Gfeed(6/23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>North America</td>
<td>2,374,878</td>
<td>73.75%</td>
<td>3,268,737</td>
<td>33.57%</td>
</tr>
<tr>
<td>Russia</td>
<td>Europe</td>
<td>184,138</td>
<td>5.72%</td>
<td>1,395,318</td>
<td>14.33%</td>
</tr>
<tr>
<td>Denmark</td>
<td>Europe</td>
<td>158,476</td>
<td>4.92%</td>
<td>609,179</td>
<td>6.26%</td>
</tr>
<tr>
<td>Germany</td>
<td>Europe</td>
<td>88,158</td>
<td>2.74%</td>
<td>602,988</td>
<td>6.19%</td>
</tr>
<tr>
<td>Canada</td>
<td>North America</td>
<td>87,242</td>
<td>2.71%</td>
<td>576,563</td>
<td>5.92%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Europe</td>
<td>65,590</td>
<td>2.04%</td>
<td>443,397</td>
<td>4.55%</td>
</tr>
<tr>
<td>Norway</td>
<td>Europe</td>
<td>39,186</td>
<td>1.22%</td>
<td>402,022</td>
<td>4.13%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Europe</td>
<td>29,815</td>
<td>0.926%</td>
<td>327,045</td>
<td>3.36%</td>
</tr>
<tr>
<td>Ukraine</td>
<td>Europe</td>
<td>19,419</td>
<td>0.603%</td>
<td>165,427</td>
<td>1.70%</td>
</tr>
<tr>
<td>Spain</td>
<td>Europe</td>
<td>16,000</td>
<td>0.497%</td>
<td>158,947</td>
<td>1.63%</td>
</tr>
<tr>
<td>Marshall Islands</td>
<td>Oceania</td>
<td>6</td>
<td>0.000186%</td>
<td>6</td>
<td>0.0000062%</td>
</tr>
<tr>
<td>Maldives</td>
<td>Asia</td>
<td>6</td>
<td>0.000186%</td>
<td>6</td>
<td>0.0000062%</td>
</tr>
<tr>
<td>Niger</td>
<td>Africa</td>
<td>6</td>
<td>0.000186%</td>
<td>6</td>
<td>0.0000062%</td>
</tr>
<tr>
<td>Suriname</td>
<td>South America</td>
<td>6</td>
<td>0.000186%</td>
<td>6</td>
<td>0.0000062%</td>
</tr>
<tr>
<td>Tonga</td>
<td>Oceania</td>
<td>6</td>
<td>0.000186%</td>
<td>6</td>
<td>0.0000062%</td>
</tr>
<tr>
<td>Samoa</td>
<td>Oceania</td>
<td>6</td>
<td>0.000186%</td>
<td>6</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Aruba</td>
<td>South America</td>
<td>8</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Dominica</td>
<td>North America</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Guyana</td>
<td>South America</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Cayman Islands</td>
<td>North America</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Mali</td>
<td>Africa</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>North America</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>San Marino</td>
<td>Europe</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Martinique</td>
<td>North America</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Burundi</td>
<td>Africa</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Comoros</td>
<td>Africa</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Togo</td>
<td>Africa</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Timor-Leste</td>
<td>Asia</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Uganda</td>
<td>Africa</td>
<td>6</td>
<td>0.000186%</td>
<td>4</td>
<td>0.0000041%</td>
</tr>
<tr>
<td>Zambia</td>
<td>Africa</td>
<td>6</td>
<td>0.000186%</td>
<td>2</td>
<td>0.000021%</td>
</tr>
<tr>
<td>Cuba</td>
<td>North America</td>
<td>6</td>
<td>0.000186%</td>
<td>2</td>
<td>0.000021%</td>
</tr>
</tbody>
</table>

of those published in the June 2023 geofeeds were geolocated to Africa. Moreover, amongst the 21 least represented countries we note that eight are located in the Americas, seven are located in Africa, and that no countries see an increase in the total geofeed IPv4 addresses geolocated to them. Assessing the regional breakdown of representation by continent as given in Table 3.3, further highlights this trend. Here we find that continents with lower proportions of industrialized countries, or of countries with strong Internet hosting infrastructure are
Table 3.3: Geofeed Geographic Representation by Continent. A breakdown of geographic representation by continent in the June 13, 2023 geofeed pull. Continents containing less countries with readily available Internet infrastructure (for example South America and Africa) are significantly less represented, or rather have substantially fewer IPv4 addresses geolocated to them within the geofeed results.

<table>
<thead>
<tr>
<th>Continent</th>
<th>Number of IPs</th>
<th>Percent of Gfeed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>4,988,562</td>
<td>45.6%</td>
</tr>
<tr>
<td>North America</td>
<td>3,852,932</td>
<td>39.6%</td>
</tr>
<tr>
<td>Asia</td>
<td>565,872</td>
<td>5.81%</td>
</tr>
<tr>
<td>Oceania</td>
<td>224,200</td>
<td>2.30%</td>
</tr>
<tr>
<td>South America</td>
<td>74,824</td>
<td>0.77%</td>
</tr>
<tr>
<td>Africa</td>
<td>29,330</td>
<td>0.30%</td>
</tr>
<tr>
<td>Antarctica</td>
<td>9</td>
<td>0.00092%</td>
</tr>
</tbody>
</table>

less represented within the geofeed results. Given geofeeds’ intended role of preempting IP ranges’ geographic mislocation by commercial IP-geolocation services, this discrepancy could further contribute to existing regional disparities in users’ Internet accessiblility and QoS if it continues to perpetuate. We discuss this in more detail in § 3.6.

3.5 Comparing Geofeed Results and Commercial IP-geolocation Estimates

To better understand the impact of uneven geofeed growth, we then compared the geofeeds with the IP-geolocation estimates from Maxmind-GeoIP2 and IPgeolocation.io, two of the commercial IP-geolocation providers’ listed in Table 3.1. In particular, we tried to determine whether we saw evidence implying that Maxmind-GeoIP2 and IPgeolocation.io consult geofeeds to inform their geolocation estimates on the IPv4 addresses reported in said geofeeds.

To do this, we periodically compared geofeed results with versions of the commercial IP-geolocation DBs pulled at roughly the same time. We find the vast majority (roughly 99%) of IPv4 addresses published with the geofeeds are present within both commercial datasets,

---

8As we note in § 3.4.2, we were only able to compare these two services due to financial/cost limitations
Table 3.4: Maxmind-Geofeed Pull-Date Pairings. Mapping of pull dates for geofeed results and matched commercial DB pulls. Pairings were selected to minimize the time between the geofeed and commercial pull dates (or vice versa).

<table>
<thead>
<tr>
<th>Gfeed Date</th>
<th>MaxmindDB Date</th>
<th>Gfeed Date</th>
<th>IPgeoloc.DB Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023-01-13</td>
<td>2023-01-13</td>
<td>2023-02-01</td>
<td>2023-01-31</td>
</tr>
<tr>
<td>2023-02-28</td>
<td>2023-02-28</td>
<td>2023-02-28</td>
<td>2023-02-28</td>
</tr>
<tr>
<td>2023-03-28</td>
<td>2023-03-28</td>
<td>2023-03-13</td>
<td>2023-03-14</td>
</tr>
<tr>
<td>2023-04-28</td>
<td>2023-04-25</td>
<td>2023-04-13</td>
<td>2023-04-11</td>
</tr>
<tr>
<td>2023-05-13</td>
<td>2023-05-16</td>
<td>2023-05-28</td>
<td>2023-05-23</td>
</tr>
<tr>
<td>2023-06-16</td>
<td>2023-06-13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and that each of the commercial services shows high levels of agreement with the geofeeds on the locations of these IPv4 addresses.

3.5.1 Comparison Methodology

To assess geofeeds’ impact on the accuracy of commercial IP-geolocation services we compared geolocation estimates from Maxmind-GeoIP2 and IPgeolocation.io to the geolocation information given in the geofeeds queried at roughly the same time.

That is, we use the geofeed data as “ground truth” and evaluate the degree to which the commercial IP-geolocation services agree with it. We show the geofeed pull dates and those of corresponding commercial DB pulls for Maxmind and IPgeolocation.io in Table 3.4. For each pairing of geofeed and contemporary commercial DB pull we performed the following steps to assess their total agreement:

1. We identified pairings of geofeed and commercial DB records where the IPv4 addresses referenced in their respective CIDR prefixes overlapped.
2. We compared the location named in each pairing’s geofeed record with the one named in the commercial DB’s estimate.\(^9\)

3. In cases where the location names did not match, we then performed reverse-geocoding on the geofeed location names to ascertain their approximate geographic coordinates.\(^10\) Since Maxmind and IPgeolocation.io both reported using the Geonames reverse geocoding database \(^2\) for this task \(^68, 87\), we decided to use it as well to maintain consistency.\(^11\)

4. Using the geofeed locations’ estimated geographic coordinates we then computed the approximate geodesic distance \(^3, 74\) between them and the (corresponding) commercial entries’ location estimates.

Using this information we then assessed the overall agreement on both the country and city levels across all IPv4 addresses that were included in both of the geofeed result and the contemporary commercial DB pull. We describe our results in § 3.5.3.

3.5.2 Geofeed vs. Commercial Comparison: Study Limitations

In addition to the limitations noted in § 3.4.2, we acknowledge that our comparison between geofeeds and commercial IP-geolocation services faces additional limitations due to the high cost of subscriptions for these services and geofeeds’ currently limited coverage of the IPv4 namespace. As a result of commercial IP-geolocation services’ high prices, we focused on

\(^9\)To account for locales having numerous names or versions of the same name (e.g. The city name for Đakovo, Croatia could also be spelled Djakovo or Dakovo), we computed the normalized Damerau-Levenshtein distance (see the pyxDamerauLevenshtein package for documentation \(^5\)) between the two location names and asserted that to match, the result had to be less 0.5.

\(^10\)While both commercial providers included the GPS coordinates of their location estimates in each record, RFC 8805 specifically notes that geofeed fields such as GPS coordinates or zip codes are deprecated and if present within a geofeed, that they should not be processed by geofeed consumers \(^77\). We therefore resort to reverse-geocoding to get a comparable estimate of the geofeeds’ named locations.

\(^11\)To once again account for many locations having multiple names or name spellings we used fuzzy matching to find many of the named locations. We found the tokenized Levenshtein distance was better suited for this task. (See the project archive \(^39\) and documentation for Python TheFuzz \(^7\) for additional details.)
two commercial providers rather than on the five services (Google, Azure, DigitalElement, Maxmind-GeoIP2, IPgeolocation.io) within the top 10 listed in Table 3.1. Since geofeeds currently account for less than 1% of the address space covered by each (respective) commercial IP-geolocation service studied, we are limited in the conclusions we can draw regarding geofeeds’ current impact on the overall accuracy of the commercial IP-geolocation services, and cannot directly compare our results to previous studies of their overall accuracy (e.g. Gharabieh et al. [84]).
Figure 3.6: Maxmind Error Distance CDF Overlay. An overlay of the error distance CDFs comparing Maxmind-GeoIP2 records to geofeed ground truth measured in January 2023 (red) and June 2023 (blue). While the proportion of matching IPv4 addresses remained at roughly 86% for the entire observation period, the mean error distance dropped from 185km in January to 169km in June.

### 3.5.3 Comparison Results

Overall, we find evidence that suggests both Maxmind-GeoIP2 and IPgeolocation.io likely consult geofeeds when updating their (respective) IP-geolocation DBs. This was characterized by very high rates at which both commercial providers provided geolocation estimates for the corresponding geofeed’s IPv4 addresses and their high accuracy rates on the geolocation of these addresses. In more detail, we found both providers covered 99% - 99.9% of their contemporary geofeeds’ IPv4 addresses. Moreover, their country and city-level accuracy for geofeed IPs roughly met or significantly surpassed their respective self-reported accuracy rates. At the country level, the observed agreement between geofeed results and commercial estimates was about 97% and 98% for Maxmind-GeoIP2 and IPgeolocation.io respectively,
Figure 3.7: Maxmind Inaccuracy Rates Compared to Geofeed Results. Maxmind’s country-level accuracy between January - June 2023 ranged from 96.1% to 97.2%. While slightly lower than their self-reported country-level accuracy [88], these steadily improved over most of the measurement period. In contrast, Maxmind’s city-level accuracy within a 5km error radius fluctuated between 88.0% and 85.1% during the same time period, but remained higher than their self-reported city-level accuracy [88].

and very close to their self-reported accuracy rates of 99% (Maxmind) and 99.9% (IPgeolocation.io) [69, 88]. City level agreement between geofeed results and commercial estimates ranged from 85.1%-86.7% (mean = 86.1%) for Maxmind and between 85.4%-91.3% (mean = 89.6%) for IPgeolocation.io. It is worth noting that both these rates of city level geofeed agreement significantly exceed the providers’ self-reported accuracy rates of 66% within a 50km radius for Maxmind-GeoIP2 and 75% for IPgeolocation.io [69, 88].

To this end, we also observe a net improvement in the mean error distance during the ob-
Figure 3.8: IPgeolocation.io Overlap with Geofeeds and Overall Accuracy. An overview of IPgeolocation.io’s rates of IPv4 address overlap and consistency with geofeeds’ geolocation over time.

3.6 Discussion

Our work highlights that while geofeeds’ coverage to date is limited, that commercial IP-geolocation services heavily rely on them as sources of ground truth geolocation data. Moreover, it indicates a potentially increasing correlation between a country’s level of industrialization and its overall representation within the geofeed results. In the case where network
Figure 3.9: IPgeolocation.io Error Distance CDF Overlay. An overlay of the error distance CDFs comparing IPgeolocation.io’s geolocation estimates to the geofeed ground truth in January 2023 (red) and on June 2023 (blue).

operator opt-in to geofeeds becomes the norm in more industrialized locations, our observations indicate that the accuracy with which commercial IP-geolocation providers would be able to locate a given IPv4 internet vantage point could become increasingly correlated with the extent to which the country housing it is industrialized. This is particularly concerning in light of existing works showing that Internet connections originating in less industrialized countries face higher rates of geoblocking [116] and that if given the option, websites/host servers will likely geofilter traffic in cases where doing so is not actually necessary [17, 79]. Since websites and online services frequently rely on commercial IP-geolocation service responses to dictate who can access them and how they behave once accessed, this would translate into a growing discrepancy in the accuracy with which web hosts would be able geolocate Internet vantage points based on their rates of industrialization. Moreover, it implies that less industrialized countries could sustain further degradation to their overall Inter-
net access and QoS as a result. For this reason, continued study of geofeeds’ expansion is necessary to identify and understand their impacts on the Internet ecosystem.

3.7 Summary

This chapter explores geofeeds’ proliferation across the IPv4 namespace and seeks to determine whether commercial IP-geolocation services rely on geofeeds as a source of ground truth geolocation information. As we describe in § 3.3, the growth of geofeed publication by network operators and their consumption by commercial IP-geolocation services is likely to have a significant impact on geoblocking. This is particularly true for CDNs that offer geoblocking, which heavily rely on commercial IP-geolocation services to support their ge-
ofiltering capabilities [4, 12, 28]. However, to date we find that network operators’ opt-in to publishing geofeeds has been limited. As such, its precise effects on geoblocking remain unclear. Despite geofeeds limited growth, we show that:

1. When geofeeds are available, commercial IP-geolocation services heavily rely on them as sources of ground truth geolocation information. As noted in § 3.5.3, this is particularly true at the country-level. This finding is especially significant as servers’ determinations of whether or not to allow inbound connection requests are frequently based on the country in which they determine the client to be based [17]. Therefore, if we assume geofeed results are accurate, the extent to which a country is represented within the geofeeds directly influences the accuracy with which commercial IP-geolocation services can recognize IP-addresses based in said country, and by extension, can discern when inbound connection requests originate there.

2. To date countries with smaller Internet presence have been underrepresented in the geofeeds. As discussed in § 3.6 this is especially concerning in light of previous research in the literature indicating that:

   - Commercial IP-geolocation services already underrepresent many countries within their DBs [99];
   - Countries with less Internet infrastructure already experience higher rates of geoblocking by websites based in regions where Internet infrastructure is more readily available (e.g. Europe and the United States) [116]; and
   - Even if unnecessary, many websites and CDN customers will opt to geoblock if given the opportunity to do so [17, 79].

To this end, Kumar et al., who study geodifferences in Android app availability, find that while these differences can stem from government takedown requests
(e.g. when an app violates a local law), the most common cause of these geodifferences is app developers restricting the geographic locations in which their apps are offered [79].

Assuming the same is true regarding website owners and their sites’ availability, the magnitude of websites unavailable in countries with less Internet infrastructure is likely to already be quite large. Moreover, should the comparative accuracy with which commercial IP-geolocation services can geolocate IPs based in these countries continue to decline, the rate at which they face geoblocking is likely to grow even larger. As such, we believe further measurement of geofeeds’ Internet coverage and their use as ground-truth data sources by commercial IP-geolocation services is not only warranted, but urgently needed to better understand the future prospects for geoblocking’s prevalence and the Internet’s accessibility around the world.
Chapter 4

An in-Depth Exploration of Smart DNS “Unblocking” Systems

Since geoblocking has become more prevalent, there has also been increased motivation to circumvent these access restrictions. Specifically, the highest ranking results from a July 2022 Google search of the term geoblocking includes websites describing how to circumvent these restrictions, as well as autocomplete query recommendations that allude to geoblock evasion such as vpn geo-blocking and how to bypass geo-blocking without vpn. Among the solutions mentioned within the top website results are commercial VPNs, Tor [34], open proxies [85, 117] and SDNS.

While not the most commonly mentioned “unblocker” within the search results, SDNS provides a stark contrast with the other tools mentioned. Unlike other tools, SDNS is geared to be lightweight and to minimize the overhead latency incurred by geoblock evasion, and has had very limited study in the open literature.

Moreover, SDNS is specifically designed for geoblock evasion and little else. As such, its setup and usage differs significantly from other previously studied unblockers. While many other geoblock evasion tools frequently require the installation of custom software, and some degree of user expertise, SDNS is simple to use and requires neither of these.

This chapter is based on the paper
Holes in the Geofence: Privacy Vulnerabilities in “Smart” DNS Services, which appeared at the 2021 Privacy Enhancing Technologies Symposium. The paper’s complete citation is given below:

Instead, a user reconfigures their computer’s DNS settings to use a DNS resolver operated by an SDNS service. The SDNS resolver “smartly” identifies resolution requests for restricted domains (hereinafter, fenced sites) and returns proxy servers’ IPs in lieu of these domains’ correct IPs. The client’s machine then directs its traffic to the specified proxy server (since that is the address to which the domains resolve), which is located within the geofence. Finally, the proxy servers relay the clients’ communication to and from these requested domains. For domains that would generally be accessible from the client’s location, DNS requests are resolved correctly. Thus, the end-user needs only browse as usual; all SDNS proxy management happens (potentially unnoticed) without additional interaction.

This chapter describes an exploration of the privacy and security properties of SDNS services—to the best of our knowledge, one of the first such studies in the open literature\(^1\). Through analyzing the architecture and behavior of deployed SDNS systems, we provide descriptions of how SDNS services operate and gain additional insight into the limitations of common geofiltering implementations. Moreover, our study of SDNS system’s security and privacy properties also highlights some of the secondary and tertiary impacts of geoblocking’s widespread use across the Internet. We note that a secondary effect – namely that geoblocked Internet users often turn to geoblock evasion tools (sometimes called unblockers) to get around geographic access restrictions – has been shown in studies of other such tools to have additional security and privacy ramifications for these individuals [75, 85, 117]. To that end, our analysis of SDNS systems also helps to give a more comprehensive understanding of these threats by focusing on those to which SDNS users and providers may be vulnerable.

Our main findings—SDNS customer IP addresses can be easily mined by third parties; SDNS substantially increases users’ vulnerability to eavesdropping; and content providers can trivially discover when users attempt to bypass their geoblocks—all threaten the privacy and/or security of SDNS customers.

\(^1\)To the best of our knowledge the first such study was performed by Fainchtein et al. [41], or the paper upon which this chapter is based.
Although SDNS may not itself be considered a privacy-preserving technology, the architectural and implementation weaknesses we describe in this chapter are relevant to SDNS users, whose use of these systems may constitute significant privacy risks.

Worse still, we observe several instances where SDNS is provided by existing VPN providers, (perhaps due to overlap in infrastructure requirements), and in some cases is advertised in a way that falsely implies that SDNS is itself a privacy-enhancing technology [15, 71, 119]. While the manner in which SDNS is marketed differs among providers, we emphasize that we found no instances in which SDNS providers describe any added privacy risks associated with using their respective services. As such, we conjecture that many SDNS users are prone to dangerous misconceptions about SDNS’s impact on their online security and privacy. (We explore these concerns in more detail in Chapter 5.)

These findings are even more concerning when we consider the rate at which SDNS services appear to be used. Specifically, SDNS does not appear to be a niche industry. As claimed by at least two SDNS providers (www.ibvpn.com and www.smartdnsproxy.com), and largely supported by our own measurements, several SDNS services appear to have over one million users each.

4.1 Background on DNS

DNS [91] is the mechanism by which hostnames are mapped to IP addresses to facilitate Internet routing. DNS is complex with several important nuances, but conceptually, DNS can be thought of as a distributed database, with mappings between hostnames and their IP addresses stored in zone files. Ordinarily, the owner of the domain (i.e., the party that registers the domain) effectively controls this mapping.

Users resolve—that is, translate a hostname to its IP address—by querying a DNS resolver. Typically, users use the resolver that is provided in the DHCP response they receive when joining a network; often, but not always, these resolvers are operated by the ISP that
provides Internet connectivity. Users also have the option of selecting a different resolver: popular choices include Google’s DNS and Cisco Umbrella’s DNS resolver.

When receiving a request, a resolver checks its cache for the queried hostname. If it finds an unexpired entry, the cached results are immediately returned. Otherwise, either the resolver returns a reference to another resolver (an iterative query) or, the resolver itself relays the request towards another resolver on behalf of the client (a recursive query) and ultimately returns the resolved IP. The resolver also caches a copy for a length of time that is defined in the corresponding zone file. The resolver that is responsible for a given domain is known as an authoritative name server and it is contacted in recursive queries when the answer is not cached by the other DNS resolvers. We found that all SDNS resolvers support only recursive queries.

While DNS supports both UDP and TCP, the former is much more common. DNS is typically neither authenticated nor encrypted. To address this and improve privacy and security, there are three main extensions to DNS that offer additional privacy features: DNSSEC, DNS-over-HTTPS (DoH) [63] and DNS-over-TLS (DoT) [65]. DNSSEC aims to ensure the authenticity of DNS data by incorporating a PKI and using signed and verifiable zone files. (Friedlander et al. provide a good overview of DNSSEC [48].) It is worth noting however, that DNSSEC only addresses the authenticity of a DNS request and not its confidentiality. That said, by ensuring response authenticity DNSSEC obviates the benefits of SDNS, and is inherently incompatible with it (since SDNS returns modified resolution results). Unlike DNSSEC, DoH and DoT both provide confidentiality of DNS requests and responses by using TLS. However, we found no SDNS providers that support either DoT or DoH.

4.2 Background on SDNS’s Proxying Behavior

As explained more fully in the next section, the proxies used by SDNS providers inspect Server Name Indication (SNI) TLS headers [22] to extract the hostname requested by the client. Once the requested hostname is obtained, the proxies simply forward TCP traffic
between the client and the destination. Such proxies are often called *SNI proxies*, and have been used as building blocks for domain fronting systems [44] (e.g., Tor’s meek [34, 42]) and more generally for proxying of Internet traffic. Using ZMap [38] scans and a novel SNI proxy testing tool, Fifield et al. identify approximately 2500 *open* SNI proxies [43] that service public requests. We find that Fifield’s list includes some SNI proxies operated by SDNS services, highlighting these services’ failure to properly authenticate requests; we explore authentication errors in more detail in §4.8.

4.3 ARCHITECTURE OF SDNS SERVICES

There are two phases of SDNS usage: registration and operation.

During the **registration** phase, a user must create an account on the SDNS service via the service’s webpage and, depending upon the service, select a payment plan. The user must also register their public-facing IP address with the SDNS service. Many services simplify IP registration by presenting the detected IP address of the user as the default and sometimes the only option. Next, the user must select a DNS resolver from a list of resolvers operated by the SDNS service. Many services advise the user to select an SDNS resolver that is geographically located nearby. This reduces the network latency incurred during DNS lookups, which, as shown in previous work [40], can significantly impact the user’s overall browsing experience. Finally, the user must configure their computer to use the selected SDNS resolver as its primary DNS resolver. All SDNS services provide detailed instructions, complete with screenshots, on how to carry out this process.

The **operation** phase is depicted in Figure 4.1. This begins when the user attempts to access geoblocked content that is supported by the SDNS service. We adopt the terminology of many of the SDNS services and refer to a geoblocked website proxied by an SDNS service as a *supported channel* or, more concisely, as a *channel*. ²

²The term is likely inspired by TV channels; SDNS effectively allows its users to “tune” to “channels” that would otherwise be unavailable.
Without loss of generality, consider a request for https://www.netflix.com, a channel supported by the user’s chosen SDNS service. The user’s DNS requests—for www.netflix.com and for domains that host web objects referenced on that page (e.g., fls.doubleclick.net)—are sent to the SDNS resolver (step 1). For each resolution request, the SDNS resolver either returns the correct IP address (e.g., via recursive lookups, as depicted in step 2) or returns the IP address of one of its proxies that reside within the requested domain’s geofence (step 3).

It is worth highlighting that SDNS depends entirely on IP-based authentication to determine whether the requesting user has completed the registration phase. If the user is not registered, the SDNS resolver’s behavior differs by SDNS provider. Most providers return a correct or non-proxy IP when resolving fenced content for non-customers. (As we show in §4.6, doing otherwise can lead to serious privacy vulnerabilities.) SDNS cannot support more robust forms of authentication since (i) DNS does not support requestor authentication and
proxies cannot rely on web-based authentication mechanisms, such as cookies; HTTPS prohibits the proxy from inspecting session cookies, since TLS encrypts all content between the client and the website.

Returning to our example of a registered SDNS user accessing geoblocked content, if the SDNS service supports https://www.netflix.com, the user’s configured device will send HTTP/S requests to the proxy IP returned by the SDNS resolver (step 4). The task of the proxy is twofold: first, it must determine which site is being requested since a single proxy may serve multiple channels (e.g., Netflix, Hulu, and ESPN). If the request is over HTTP, then the proxy can inspect the Host HTTP header, which is mandatory in HTTP/1.1. For encrypted HTTPS traffic, SDNS exploits TLS’ Server Name Indication (SNI) extension [22] that allows the requested site to be communicated as cleartext. SNI is intended to allow a single server to host multiple domains and serve the correct TLS certificate during the exchange. SNI is a popular TLS extension and has been found to be present in 99% of TLS connections [49]. In the context of SDNS, the use of SNI allows the SDNS proxy to interlope on the exchange and learn to which domain (e.g., Netflix) it should send proxy traffic.

Second, the proxy must actually forward the traffic (step 5). SDNS proxies operate transparently and function as TCP endpoints for both the requesting client where it poses as the web server and the web server where it poses as the client. SDNS proxies merely relay data received through one TCP connection to the other, and vice versa; doing otherwise would disrupt TLS (HTTPS) traffic between the client and the server.

The SDNS service does not necessarily have to proxy all web objects that are included in a requested webpage (step 6). For example, the SmartDNSProxy provider does not proxy requests to fls.doubleclick.net, even though such web objects are referenced on www.netflix.com. This has the advantage that it decreases the proxy’s workload and operating cost.
4.3.1 DNSSEC and Encrypted SNI

SDNS services are entirely incompatible with DNSSEC, since the latter provides origin authentication of DNS records. Of course, SDNS resolvers do not support the DNSSEC extensions, making this incompatibility moot until browsers and/or operating systems begin to require DNSSEC support.

Cloudflare co-introduced and adopted \[53\] encrypted SNI \[104\], which eliminates a privacy weakness of SNI by encrypting the requested hostname between the client and the receiving web server. Encrypted SNI would thwart the current SDNS architecture by hindering a proxy’s ability to identify the site being requested. However, as of July 2021, encrypted SNI, and its successor encrypted client hello \[105\] is either unsupported or not enabled by default in the latest release versions of Chrome, Firefox, Safari, Brave, and Microsoft Edge.

4.3.2 Contrasting with VPN Services

A stark difference between SDNS services and VPNs is that the former has no obvious on/off mechanism. VPNs require starting an application and authenticating to the VPN provider. In most settings, the VPN is not on by default, and there are visual indicators on the desktop when the VPN is in use. User intervention is also required to reestablish the connection to the VPN after machine reboots. That is, the use of the VPN requires intentional actions by the user.

In contrast, while SDNS providers offer their customers some helpful instructions and tools to configure their DNS settings, SDNS is much less user-friendly with respect to activation or deactivation. Aside from the availability of certain video streaming services there are no obvious indicators that the SDNS service is in use. Due to this opacity in SDNS services, we posit that many SDNS users will forget the current status of their DNS settings and effectively always be performing DNS lookups through the SDNS’ resolvers. We discuss the security and privacy implications of continuously using SDNS services in §4.6. SDNS is also unlike VPNs in that SDNS does not encrypt content between the user’s device and the
proxy. It is unable to do so, since its use is entirely invisible to the user’s computer. For non-
HTTPS traffic, SDNS increases the attack surface by allowing any potential eavesdropper
between the client and the proxy to perform man-in-the-middle manipulation. In the case
of HTTPS traffic, an eavesdropper can inspect SNI headers to learn which domain names
are being requested.

4.3.3 SDNS Marketplace

To understand the SDNS marketplace, we performed simple Google queries to identify po-
tential providers. We found that many SDNS providers are also VPN providers, advertising
SDNS alongside VPN services and usually at a lower cost\(^3\). SDNS, unlike VPNs, is not a
privacy enhancing technology, but the commingling may confuse users about the security
and privacy properties of SDNS. In at least two instances (ibVPN and VPNUK), SDNS
providers marketed SDNS alongside their VPNs as privacy-enhancing services.

In total, we identified 25 SDNS providers. Using information available on their webpages,
we catalogued (i) their prices and subscription plan offerings, (ii) the IP addresses of their
DNS resolvers, and (iii) the countries in which the providers appeared to be registered.

The names, monthly costs, and locations of the 25 identified SDNS providers are listed
in Table 4.1, including 15 providers which we focused on as part of our analysis. (These
providers were selected based primarily on their search rank when querying Google for SDNS
providers and their costs.) The 25 SDNS providers spanned 12 countries, where the country
of origin was determined by searching for contact information (i.e., mailing addresses) listed
on the providers’ web pages. When searching for listed contact addresses, we also noticed
that a number of the Turkish SDNS providers mention on their respective webpages that
they belong to a single parent company.

\(^3\)In fact, as we note in Chapter 5, most SDNS providers offer both services and sell them as a
single service bundle.
**Company Aliasing**  During our analysis of SDNS services, we gathered evidence that strongly suggests that some of the SDNS providers are in fact the same company advertising under multiple name brands.

We identify numerous instances in which SDNS providers share infrastructure. To do so, we determine a potentially incomplete set of proxies used by each SDNS service by querying their DNS resolvers for supported channels (e.g., Netflix), and then comparing the returned IP addresses with a large ground-truth dataset, which was obtained by resolving the hostnames from a distributed network of RIPE Atlas nodes [107]. Our methodology for detecting shared proxies is described in more detail in § 4.4.

We find that the SmartDNSProxy, Trickbyte, and Uflix SDNS services share extensive infrastructure; specifically, we identify 10 proxies that are used by all three providers, seven that are shared between SmartDNSProxy and Trickbyte, and 14 shared proxies between Uflix and SmartDNSProxy. We additionally note that both Trickbyte and SmartDNSProxy’s websites are served from the same /16 network, previously shared TLS certificate subjects, and are registered using the same domain registrar.

Upon additional inspection, we also discover evidence implying that CactusVPN and SmartyDNS are likely operated by a single entity. Specifically, these two providers share at least four proxies, and exhibit similar proxying behavior patterns, which we describe in more detail in §4.8.

The rationale for operating multiple seemingly distinct SDNS services in this manner is unclear. We conjecture that such a strategy may attract more customers, since there are several sites that feature reviews and rankings of SDNS services. Operating as multiple services increases the chances of appearing at the top of at least some rankings. This is similar to findings that multiple VPN services may be operated by the same entity [75], perhaps also to gain advantage in VPN review and ranking sites.

---

4.4 Identifying SDNS Proxies

To bypass geoblocking restrictions, SDNS services rely on a network of strategically placed proxies and DNS resolvers. As part of our analysis and to better understand these services’ ecosystem, we attempted to answer the questions: *where are SDNS resolvers and proxies located?; who hosts them?; and what was the likely motivation behind these choices?*

As a first step, to better understand to whom the SDNS providers are catering their services, we determine the geographic location of the SDNS providers’ DNS resolvers.

To reduce network latencies incurred during DNS resolutions, SDNS providers often recommend that their customers select an SDNS resolver that is in close physical proximity. Understanding where SDNS providers place their resolvers thus provides hints as to where they envision the best opportunities for attracting customers. Using MaxMind’s geolocation service, we map the listed DNS resolvers for each proxy to a location, and report the countries with the most resolvers in Table 4.2. As a point of comparison, we first note the locations where SDNS providers are incorporated in Table 4.1. With the exception of the United States, we note that the locations of potential customers mostly differ considerably from the locations in which the SDNS providers operate. In short, it appears that SDNS providers mostly tailor their services to international customers.

While the locations of the resolvers indicate potential customer markets, the proxies’ locations correlate to regions that lay within supported channels’ geofences. However, due to the nature of the modern web and the prevalence of content distribution networks (CDNs), identifying SDNS providers’ proxy servers proved complicated.

At a high level, the task entails querying an SDNS resolver for a hostname and identifying whether the returned IP address was (i) accurate or (ii) that of a proxy server. In reality, unfortunately, DNS resolution is fairly complex. Rather than consistently returning a single IP, multiple queries to a single domain name on a normal (non-SDNS) resolver return the IP of the host that can serve the website’s content fastest, given the current network state and the requester’s network location. When accounting for the widespread use of CDNs, this
also means seemingly unrelated sites can be resolved to the same IP address (since multiple sites can share CDN replicas). In short, it is difficult to enumerate all possible valid IP addresses that belong to a given hostname; this is especially true of popular sites including the channels supported by SDNS providers, since such sites tend to heavily rely on CDN services.

We identify proxy IPs using a two-phase approach: at a high level, we first identify a set of candidate IPs we believe may be SDNS proxies, and then verify them. To generate candidate IPs, we queried 10 SDNS providers’ resolvers for two sets of domains: the Alexa top 1,000 most popular sites [18], and the hostnames of the channels advertised as being supported by the SDNS provider. We then compared the returned list of IPs against a ground truth dataset generated by making over 32,000 DNS requests to Google’s and CloudFlare’s DNS resolvers, as well as requests from RIPE Atlas probes to their local resolvers. The latter was included to increase the geographic and network diversity of the requesting DNS clients. The ground truth dataset was constructed using DNS queries conducted between February 14 and March 31, 2019, and again between April 25 and May 3, 2019. Finally, we generated a candidate list of hostname-to-resolved-IP pairings by first considering the responses from SDNS resolvers and then eliminating entries for which an IP in the same /24 appeared in the ground truth dataset.

To verify the candidate IPs as proxies, we attempted to fetch content via a candidate proxy from both a machine that was registered with the SDNS service and one that was not. Conceptually, if the candidate IP is not a proxy and is a legitimate IP address that serves content for the site, it should serve the content regardless of the requestor’s IP; on the other hand, if it is a proxy, then the proxy should only serve content for the IP that is associated with one of its customers. That is, we expect actual proxies to serve requests that originate from a registered IP, but to deny the same content requests from IPs that are not associated with the SDNS provider.
Using two machines, one whose IP we registered with the SDNS service, and one whose IP was not registered, we sent well-formed HTTP/S requests with the Host HTTP and SNI headers properly set to the candidate proxy using curl, and compared the results. We confirm a candidate IP as an actual proxy if and only if the HTTP/S request from the registered machine was successful (i.e., resulted in a 200 OK HTTP response) and the request from the non-registered machine was not.

Overall, we were able to definitively identify 54 distinct proxy IPs across five of the evaluated SDNS providers.

Table 4.4 lists the most common countries where proxy servers are located. We note that the most popular locations—the United States, the United Kingdom, and India—are also the nations that host a large fraction of the channels offered by SDNS providers. This suggests that proxies are indeed placed close to content providers.

4.5 SYSTEM AND THREAT MODELS

There are several actors in the SDNS ecosystem: SDNS providers sell geoblock-evasion services to customers in order to provide more unfettered access to geoblocked content providers. (We also refer to customers as users.) The SDNS infrastructure is composed of one or more provider-operated resolvers and one or more proxies. Additionally, the SDNS resolvers depend on the traditional DNS infrastructure, since customers’ DNS queries correspond not just to supported content providers (e.g., www.netflix.com), but also to unproxied domains (e.g., petsymposium.org).

This chapter explores the privacy and security implications of SDNS to both customers and SDNS providers, and thus we consider two separate threat models:
Table 4.1: Identified SDNS Providers. As indicated by their names, many double as purveyors of commercial VPN access.

<table>
<thead>
<tr>
<th>Provider</th>
<th>Monthly Cost</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>AceVPN</td>
<td>$5.95°</td>
<td>USA</td>
</tr>
<tr>
<td>Blockless</td>
<td>$3.32°</td>
<td>Canada</td>
</tr>
<tr>
<td>*BulletVPN</td>
<td>$10.98°</td>
<td>Estonia</td>
</tr>
<tr>
<td>*CactusVPN</td>
<td>$4.99</td>
<td>Moldova</td>
</tr>
<tr>
<td>*DNSFlex</td>
<td>$5.00</td>
<td>Canada</td>
</tr>
<tr>
<td>*DNSTrick</td>
<td>$4.95</td>
<td>Unknown</td>
</tr>
<tr>
<td>*GetFlix</td>
<td>$39.00°</td>
<td>Turkey</td>
</tr>
<tr>
<td>*HideIPVPN</td>
<td>$4.95</td>
<td>Unknown</td>
</tr>
<tr>
<td>Hide-my-IP</td>
<td>$4.95</td>
<td>USA</td>
</tr>
<tr>
<td>*ibVPN</td>
<td>$10.95</td>
<td>Romania</td>
</tr>
<tr>
<td>Ironsocket</td>
<td>$4.16</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>*Keenow</td>
<td>$5.79</td>
<td>Israel</td>
</tr>
<tr>
<td>Le-VPN</td>
<td>$9.95°</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>*Overplay</td>
<td>$4.99</td>
<td>USA</td>
</tr>
<tr>
<td>simpletelly</td>
<td>$4.99</td>
<td>Turkey</td>
</tr>
<tr>
<td>*SmartDNSproxy</td>
<td>$4.90</td>
<td>Turkey</td>
</tr>
<tr>
<td>*SmartyDNS</td>
<td>$4.90</td>
<td>Moldova</td>
</tr>
<tr>
<td>StrongDNS</td>
<td>$5.00</td>
<td>USA</td>
</tr>
<tr>
<td>*TrickByte</td>
<td>$2.99</td>
<td>Turkey</td>
</tr>
<tr>
<td>TVWhenAway</td>
<td>£7.99</td>
<td>UK</td>
</tr>
<tr>
<td>*Uflix</td>
<td>$4.90°</td>
<td>Turkey</td>
</tr>
<tr>
<td>Unblock-us</td>
<td>$4.99</td>
<td>Cyprus</td>
</tr>
<tr>
<td>*Unlocator</td>
<td>$4.95</td>
<td>Denmark</td>
</tr>
<tr>
<td>VPNSecure</td>
<td>$9.95</td>
<td>Australia</td>
</tr>
<tr>
<td>*VPNUK</td>
<td>£5.99</td>
<td>UK°</td>
</tr>
</tbody>
</table>

* indicates a provider included in our measurement analysis  
• indicates a lifetime cost  
° indicates a cost that also includes VPN services  
°° contact info in UK, but company registered in Belize
Table 4.2: Top Ten Countries with the Most SDNS Resolvers.

<table>
<thead>
<tr>
<th>Country</th>
<th>N. Resolvers</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>9</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>4</td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
</tr>
<tr>
<td>Australia</td>
<td>3</td>
</tr>
<tr>
<td>Germany</td>
<td>3</td>
</tr>
<tr>
<td>India</td>
<td>3</td>
</tr>
<tr>
<td>Netherlands</td>
<td>3</td>
</tr>
<tr>
<td>Denmark</td>
<td>2</td>
</tr>
<tr>
<td>South Africa</td>
<td>2</td>
</tr>
<tr>
<td>Singapore</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.3: Top ASes with the Most SDNS Resolvers.

<table>
<thead>
<tr>
<th>AS Name &amp; Num.</th>
<th>N. Resolvers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon (16509)</td>
<td>21</td>
</tr>
<tr>
<td>DigitalOcean (14061)</td>
<td>15</td>
</tr>
<tr>
<td>SoftLayer Techs (36351)</td>
<td>10</td>
</tr>
<tr>
<td>Choopa LLC (20473)</td>
<td>5</td>
</tr>
<tr>
<td>Iomart (20860)</td>
<td>5</td>
</tr>
<tr>
<td>Linode LLC (63949)</td>
<td>3</td>
</tr>
<tr>
<td>OVH SAS (16276)</td>
<td>3</td>
</tr>
<tr>
<td>SiteHost NZ (45179)</td>
<td>3</td>
</tr>
<tr>
<td>ASERGO (30736)</td>
<td>2</td>
</tr>
<tr>
<td>Datacamp Ltd (60068)</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.4: Top Ten Countries with the Most SDNS Proxies.

<table>
<thead>
<tr>
<th>Country</th>
<th>N. Proxies</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>15</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>13</td>
</tr>
<tr>
<td>India</td>
<td>6</td>
</tr>
<tr>
<td>Australia</td>
<td>3</td>
</tr>
<tr>
<td>Denmark</td>
<td>3</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
</tr>
<tr>
<td>France</td>
<td>2</td>
</tr>
<tr>
<td>Canada</td>
<td>2</td>
</tr>
<tr>
<td>Germany</td>
<td>2</td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.5: Top ASes with the Most SDNS Proxies.

<table>
<thead>
<tr>
<th>AS Name &amp; Num.</th>
<th>N. Proxies</th>
</tr>
</thead>
<tbody>
<tr>
<td>DigitalOcean (14061)</td>
<td>8</td>
</tr>
<tr>
<td>QuadraNet (8100)</td>
<td>6</td>
</tr>
<tr>
<td>Iomart (20860)</td>
<td>4</td>
</tr>
<tr>
<td>Level 3 Parent LLC(3356)</td>
<td>4</td>
</tr>
<tr>
<td>ASERGO (30736)</td>
<td>3</td>
</tr>
<tr>
<td>OVH SAS (16276)</td>
<td>2</td>
</tr>
<tr>
<td>GleSYS AB (42708)</td>
<td>2</td>
</tr>
<tr>
<td>Compuweb (51905)</td>
<td>2</td>
</tr>
<tr>
<td>Melbikomas UAB (56630)</td>
<td>1</td>
</tr>
</tbody>
</table>

57
Table 4.6: Summary of Attacks. The adversary, required adversary capabilities, and the target of the attack are listed for each attack.

<table>
<thead>
<tr>
<th>Vulnerability</th>
<th>Adversary</th>
<th>Required Adversary Cap.</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>§4.6.1: Enumerating customers (by IP)</td>
<td>Internet user</td>
<td>reg. domain name; spoof UDP</td>
<td>Customer Threat</td>
</tr>
<tr>
<td>§4.6.2: Real-time SDNS customer identification</td>
<td>Content provider</td>
<td>operate website; view web logs</td>
<td>Customer Threat</td>
</tr>
<tr>
<td>§4.6.2: Real-time proxy server discovery</td>
<td>Content provider</td>
<td>operate website; view web logs</td>
<td>SDNS Provider Threat</td>
</tr>
<tr>
<td>§4.7: Increased risk of traffic analysis</td>
<td>Network eaves.</td>
<td>observe DNS or proxy traffic</td>
<td>Customer Threat</td>
</tr>
<tr>
<td>§4.8: Payment bypassing / free use of pay service</td>
<td>Internet user</td>
<td>send DNS resolution requests</td>
<td>SDNS Provider Threat</td>
</tr>
<tr>
<td>§4.9: Exposure to analytics / business analysis</td>
<td>Internet user</td>
<td>send DNS resolution requests</td>
<td>SDNS Provider Threat</td>
</tr>
</tbody>
</table>
4.5.1 Customer Threats

This chapter considers attacks on customer privacy that expose a customer’s IP address, either to the content provider or to any outside party. We note that such exposure could potentially present legal risks to SDNS users\(^5\), or result in users being banned by content providers.

It is worth emphasizing that, as with other work (cf. [34, 121]), we consider IP addresses to be sensitive information. Indeed, the EU’s General Data Protection Regulation (GDPR) and the California Privacy Protection Act of 2018 both consider a user’s IP address to be personally identifiable information under certain circumstances [29, 31]. We describe several means by which IP addresses can be mapped to specific individuals in Chapter 2, but highlight here that IP-to-individual mappings are commercially available (e.g., from Experian) and that IP-to-individual search engines are also available (e.g., https://thatsthem.com/reverse-ip-lookup).

Additionally, we argue that the exposure of customer IP addresses to any arbitrary outside party falls well outside of the norms that users expect from their Internet services. We can think of no other example in which a service allows outside parties to enumerate all of its users. Perhaps more importantly, we did not find any SDNS service that advises its customers about any such exposure.

Finally, we consider customers’ increased vulnerability to traffic analysis due to the use of SDNS. We consider the additional risk not just to traffic directed towards content providers, which would take longer paths due to proxying, but also other more general Internet traffic that users may not expect to be proxied.

\(^5\)In the U.S., the use of SDNS may technically violate the Computer Fraud and Abuse Act, which criminalizes “exceed[ing] authorized access” of a computer system and imposes civil liability on the perpetrator [118].
4.5.2 Provider Threats

We also explore the privacy and security risks of operating an SDNS service. These consists of vulnerabilities that either (1) harm the operation of the SDNS provider or (2) reveal potentially sensitive information about its operation.

More concretely, such threats include fundamental weaknesses in the SDNS architecture that allow a content provider to detect (and thus block), in real-time, the use of SDNS. (We note that this is a more powerful attack than attempts to enumerate SDNS proxies, since it avoids an arms race between discovering proxies and spinning up new proxies.) Additionally, we consider to be in-scope attacks that target the financial operation of an SDNS service and allow users to bypass payment and effectively access the service for free.

Finally, our threat model includes the exposure to analytics that enables outsiders to perform competitive analysis on the SDNS provider. This includes the ability of a third-party to estimate the number of users and revenue of an SDNS service, as well as to gauge the relative popularity of the channels that it proxies.

4.5.3 Adversaries and Adversarial Capabilities

We consider several adversaries that pose threats to either customers or providers. Our content provider adversary operates a channel that is targeted for geoblock bypass by an SDNS provider. The content provider has the ability to modify its website and inspect its own web logs.

The network eavesdropper adversary is a passive network observer. We consider two variants of our network eavesdropper: an eavesdropper who is located between the client and the client’s SDNS resolver, and a network eavesdropper that is located between the SDNS proxy and the geoblocked destination website (see §4.7). An example of the former is the SDNS user’s ISP; an example of the latter is a government or AS that monitors or hosts the proxy server. The network eavesdropper can inspect intercepted packets. For the eavesdropper who observes DNS traffic, our attack is effective when the DNS request and
response are not encrypted. We are not aware of any SDNS service that supports encrypted DNS resolution (i.e., with either DoT [65] or DoH [63]).

We also present a number of attacks that can be carried out by nearly any arbitrary third-party Internet user; we call this adversary the **Internet user adversary**. The Internet user adversary can exploit the authentication and authorization failures we identify in §4.8 to use SDNS services without having to pay for them. We show that such attacks are possible and can be carried out by any Internet user.

An Internet user adversary can also probe the caches of SDNS providers’ DNS resolvers to infer information about the popularity of the SDNS providers as well as which channels are most often used by the providers’ customers (see §4.9). Here, we require that our Internet user adversary be able to identify the DNS resolvers used by an SDNS provider. We note that resolvers are listed on SDNS providers’ websites since SDNS customers need such information to configure their computers to use SDNS.

Finally, the Internet user adversary can carry out the customer enumeration attack (see §4.6.1). Here, three additional capabilities are needed: the adversary needs to be able to (1) register a domain name of the adversary’s choosing, (2) operate the authoritative name server for that domain, and (3) be capable of sending spoofed UDP packets.

We summarize our main security and privacy findings in Table 4.6. We note that our threat models exclude geoblock circumvention. Although this may be reasonably considered a security threat to content providers since it bypasses an authentication check, this is the intended function of SDNS services. Our focus in this chapter, rather, is to shed a light on the previously undocumented privacy and security risks of SDNS.

### 4.6 Privacy Vulnerabilities in SDNS Designs and Implementations

SDNS’ architecture and implementations lead to several privacy and security risks, which we describe below. The relevant threat models for each vulnerability defined in §4.5, is listed in Table 4.6.
4.6.1 Client Enumeration Attacks

Standard DNS does not support client authentication, and hence SDNS providers must rely on IP-based authentication to identify customers. This is noteworthy as all tested SDNS services were found to primarily support UDP-based DNS. The use of IP-based authentication, coupled with the ease at which UDP-based DNS requests can be forged leads to serious privacy vulnerabilities.

We discovered architectural weaknesses in two SDNS services that allow a third-party attacker (the Internet user adversary described in §4.5) to query the SDNS service and, in so doing, determine whether a target IP address belongs to one of its registered customers. When repeated, this attack allows the attacker to enumerate the IP addresses of these services’ customers. For ease of exposition, we refer to an IP address associated with a customer of the SDNS provider as being a registered IP. The attack requires no client interaction and will reliably reveal whether an IP address is registered even if the customer is not actively using the SDNS service, or even if they are not currently online.

As discussed in §4.5, an adversary who learns the IP addresses of SDNS users could potentially also combine this information with existing IP-to-individual mappings to determine the users’ identities. This, in turn, could enable targeted cease and desist notifications. Even without resolving particular identities, knowledge of SDNS users’ IP addresses is sufficient to deliver abuse notifications to the operators of the users’ networks (e.g., their ISPs), akin to how movie and music trade associations communicate their perceived violations of the U.S. DMCA.

The client enumeration attack requires only that the adversary (i) registers a domain name of the adversary’s choosing and (ii) operates its own authoritative domain server for that domain. The adversary can be located far from the SDNS service and need not intercept any messages destined to SDNS resolvers. While the attacker can be any Internet user who meets the above criteria, we posit that content providers, trade associations, and content
producers (or their copyright holders) might be especially motivated to enumerate the users of SDNS.

The attacker exploits a specific SDNS behavior in which the service’s DNS resolvers send distinct responses to customers’ and non-customers’ requests. At a high level, the adversary uses these two different behaviors to deduce whether an arbitrary IP address is a customer of the service or not. This process can then be repeated for all IPv4 addresses or a target set of IP addresses for which the attacker is interested.

Figure 4.2 presents an overview of our attack. To determine whether an arbitrary IP address, say 1.2.3.4, is registered, an attacker who controls the domain attackerdomain.com forges an otherwise well-formed DNS request to the SDNS resolver for nonce.attackerdomain.com, purportedly originating from 1.2.3.4 (step 1 in Figure 4.2, left), where nonce is a unique identifier. If 1.2.3.4 is a registered IP address, the forged query for nonce.attackerdomain.com would cause the SDNS resolver to correctly resolve nonce.attackerdomain.com to its IP address (step 2, left) via recursive DNS lookups. This would be the case, as the hostname nonce.attackerdomain.com does not correspond to any channel supported by the service.

We emphasize that to support general web browsing, SDNS resolvers must correctly resolve hostnames for domain names they do not support. Additionally, the use of a unique nonce prevents nonce.attackerdomain.com from being cached at the resolver. This ensures that the request is propagated to the authoritative name server for attackerdomain.com, where it can be observed by the adversary. Finally, the IP address for nonce.attackerdomain.com or for an error if it is not found, is relayed back to the SDNS provider’s resolver (step 3, left) and forwarded onto the forged IP address X (step 4, left), where it is likely discarded.

The right-side of Figure 4.2 shows the alternative case in which the IP address 5.6.7.8 is tested and is not registered with the SDNS provider. Here, we rely on a particular behavior of certain SDNS providers; namely, that they do not resolve requests from non-customers.
Figure 4.2: SDNS IP Enumeration Attack. The two possibilities for the client enumeration attack: either the candidate IP address belongs to an SDNS customer (top) or not (bottom).
We found two slight variations of susceptible behavior. The ibVPN SDNS service responds to non-customer hostname resolution requests by returning a fixed IP address belonging to a website it operates; the website redirects the user to an error page. This scenario is depicted in step 2 in Figure 4.2, right. In contrast, the VPNUK service does not respond at all to non-customer DNS resolution requests.

Both behaviors allow the attacker to determine whether an arbitrary IP address, in this case 5.6.7.8, is a customer: if it is, it will receive the recursive lookup request from the SDNS resolver and can observe this request at its authoritative name server; if 5.6.7.8 is not a customer, the request will not appear.

To validate our attack, we performed a proof-of-concept experiment using the ibVPN and VPNUK SDNS providers. We discuss the ethics of our experiment in §4.12. Our procedure was identical for both systems: we purchased an account on the system and registered our client IP address. We also purchased a domain name and configured an authoritative name server hosted at Georgetown University for that domain. We confirmed that requests originating from our client’s IP to resolve a unique subdomain were recursively resolved and observed at our authoritative name server.

Next, we confirmed that requests sent from an unregistered IP (also operated by us), either yielded false static responses (ibVPN) or no responses at all (VPNUK); in both cases, the requests originating from the other, non-registered IP address did not propagate back to our authoritative name server.

Finally, to complete the attack, we acted as the attacker using a third IP on a different network. The attacker forged two requests: one purportedly from the registered IP and one from the non-registered IP. We confirmed that only the forged requests that purported to be from the registered IP address were relayed to our authoritative name server.

**IPv4-space Enumeration** We used the ZDNS tool from the ZMap Project [38] to estimate the service capacity of our institution’s DNS server. ZDNS performs highly parallel
DNS lookups using lightweight Go threads, and is useful for efficiently resolving a large num-
ber of domains against a DNS resolver. We emphasize that we did not use ZDNS against the
ibVPN or VPNUK resolvers since ZDNS could potentially disrupt their services. We use the
performance measurements of our institution’s DNS resolver only to form a rough estimate
of the length of time it would require to enumerate all $2^{32}$ potential IPv4 addresses.

We find that Georgetown’s DNS resolvers can resolve the top 10,000 Alexa sites in
7.462s (1340.12 requests/second), while consuming less than 1 MBps. At this rate, it would
require approximately 5.3 weeks of continuous queries to enumerate all possible customer
IPs (again, under the assumption that the SDNS provider’s resolver has comparable perfor-
mance). While such a sustained rate of access is likely unrealistic, we note that large ISPs
can be fully enumerated in under a day (e.g., Comcast has approximately 71 million IP
addresses [70]).

Mitigations Our attack relies on SDNS resolvers that resolve an attacker-controlled do-
main only when the purported requester is a customer. The attack can be partially mitigated
by consistently and correctly resolving domains for all hostnames that are not associated
with a supported channel. Indeed, one day after we disclosed our attack to ibVPN, the
ibVPN service implemented this mitigation.

We emphasize that although this fix disallows arbitrary third-parties from enumerating
customers, it will not prevent the operators of a supported channel (e.g., Netflix) from
carrying out the attack. For example, the content operator can register subdomains (e.g.,
nonce.netflix.com) and forge a DNS request from a candidate IP X to determine if X is
associated with a customer of the SDNS service. The SDNS provider cannot apply the
above fix here, since to route around geo-filters, it needs to respond to the client with
an incorrect resolution containing the IP address of a proxy when the requested site is a
supported channel.

A more robust mitigation is for the SDNS resolver to accurately resolve all resolution
requests. When the requested hostname corresponds to a supported channel, the SDNS
resolver can ignore the correct IP address and instead return the address of its proxy to the requesting client. However, while this fully mitigates the attack, it also allows a content provider (i.e., channel operator) with knowledge of the SDNS service to precisely measure how often the SDNS service is being applied to bypass its geofilter: it can inspect its own authoritative name server’s logs for relayed requests from the SDNS resolver.

4.6.2 De-proxying by the Content Provider

It is relatively straightforward for a geoblocking content provider (i.e., a website operator) to (i) detect and prevent access from an SDNS service and (ii) identify the true IP address of the SDNS customer. A de-proxying attack requires the content provider to insert content into its web page that does not require a DNS lookup. By causing DNS resolution to be skipped, the content provider prevents the use of a proxy and forces the client to perform a direct access.

Without loss of generality, consider a content provider istreamvideos.net, where istreamvideos.net resolves to the IP address 1.2.3.4. To perform a de-proxying attack, the content provider serves the partial content `<IMG src="https://1.2.3.4/image.-jpg?session_id">` where session_id is a unique tag that can link the web requests to istreamvideos.net with those to 1.2.3.4. The client’s browser will process the above IMG tag and directly access the image at 1.2.3.4 since the unused SDNS resolver loses the opportunity to return the IP of a proxy. The content provider then checks whether the two linked requests originated from the same IP; significantly differing requesting IP addresses (e.g., from different autonomous systems) indicates the use of an SDNS service.

As a proof-of-concept, we performed a de-proxying attack against ourselves, using the Hide-My-IP SDNS service. Hide-My-IP proxies all connections: its DNS resolver returns the IP address of a single proxy regardless of the requested domain. When the proxy receives HTTP requests from the client, it inspects the HOST HTTP header or the SNI TLS header.

---

6Although it is not particularly common (or advised), some certificate authorities (e.g., Global-Sign [54]) will issue IP-based certificates.
to identify the requested destination, and then acts as a transparent TCP proxy. Since Hide-
My-IP proxies all sites, we can trivially become a “content provider” by simply instantiating
a web server. As described above, we constructed a simple web page that included an IMG
tag whose source (“src”) was specified by IP rather than hostname. We confirmed that our
web server logs revealed that the domain-based request for the webpage had a different
requesting client than the one for the IP-specified image; the former showed the proxy IP
address and the latter revealed the client’s IP address.

The deproxying attack enables a content provider to learn which of its users use SDNS.
Unlike the client enumeration attacks described in §4.6.1, the deproxying attack may directly
implicate a user of the content provider if the provider requires users to first log in before
accessing content. It also provides a real-time mechanism for immediately detecting the use of
SDNS, and is thus a far more practical means of protecting against geoblock circumvention
than frequently enumerating all SDNS users. Once an SDNS user has been identified, the
content provider could either suspend or terminate that user’s account, or simply disallow
use of the service while SDNS is in use.

There are no clear mitigations to a de-proxying attack, and moreover, de-proxying attacks
are particularly worrisome for users who misunderstand the privacy properties of SDNS
services. While SDNS services do not advertise anonymity, end-users could be confused
about the kinds of protections (or lack thereof) that these services provide, especially when
these same providers sell VPN services as their primary offering. This confusion could put
end-users in restrictive regimes at particular risk, if they access censored content with an
expectation that their accesses are anonymous.

The deproxying attack also presents a threat to SDNS services. We found instances
in which content providers blocked access from a handful but not all SDNS proxies. This
indicates a “whack-a-mole” defense in which content providers attempt to identify and block
proxies. This arms race generally works in the SDNS provider’s favor, since cloud-hosted
proxies can easily change IP addresses. (This same whack-a-mole strategy is also used to
find VPN services’ egress points.) The deproxying attack avoids this arms race by identifying in real-time the use of SDNS, and thus enabling immediate discovery of SDNS proxy servers as soon as they are utilized. In short, the content provider can apply this attack to entirely eliminate the utility gained by using an SDNS service.

4.7 Susceptibility to Eavesdropping

SDNS services increase their users’ susceptibility to eavesdropping. We explore this increased risk across two dimensions: eavesdropping on DNS requests and eavesdropping on proxies.

4.7.1 Eavesdropping on DNS Requests

A log of DNS queries provides a fairly complete record of which sites and services were accessed by a requestor. SDNS customers configure their computers to send all DNS queries to SDNS resolvers, regardless of whether the queries pertain to fenced or unfenced websites. This provides SDNS services with a comprehensive set of potentially sensitive metadata about their customers. We emphasize that this is in stark contrast to using VPNs, whose use can be easily toggled on and off and whose active use is typically indicated by visible cues presented to the user. That is, the “set and forget” configurability of SDNS services, described in §4.3, has important implications to users’ privacy.

Longer Paths Increase Susceptibility to Eavesdropping  The architecture of SDNS services risks exposing their users’ Internet metadata to third parties, beyond the SDNS provider. DNS requests and responses are usually (and, in the case of SDNS, we believe always) sent unencrypted\(^7\), allowing eavesdroppers between the client and the resolver to learn which hostnames are being requested, and by whom.

For regular or non-SDNS DNS resolution, the resolver is typically located near the requestor, and is often operated by the requestor’s ISP, which we note, learns the sites being

\(^7\)The Firefox web browser now uses encrypted DoT [65] to Cloudflare’s DNS resolvers by default. However, SDNS users would need to disable this setting.
Table 4.7: Mean Number of ASes Traversed per DNS Network Path (by Region).

Average number of ASes encountered in network paths from various geographic regions to (1) Cloudflare’s and Google’s DNS resolvers (“Public Resolver”) and (2) 108 SDNS resolvers (“SDNS Resolver”). Percentage increases (relative to the public resolvers) are shown in parentheses.

<table>
<thead>
<tr>
<th>Client Location</th>
<th>Public Resolver</th>
<th>SDNS Resolver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1.50</td>
<td>2.74 (82.67%↑)</td>
</tr>
<tr>
<td>Belgium</td>
<td>1.00</td>
<td>2.60 (160.00%↑)</td>
</tr>
<tr>
<td>Brazil</td>
<td>2.00</td>
<td>2.31 (15.50%↑)</td>
</tr>
<tr>
<td>Japan</td>
<td>2.00</td>
<td>2.26 (13.00%↑)</td>
</tr>
<tr>
<td>United States</td>
<td>2.00</td>
<td>3.10 (55.00%↑)</td>
</tr>
</tbody>
</table>

requested by virtue of forwarding their traffic. That is, using a local resolver poses little additional privacy risk. Public DNS services, such as those offered by Google, Cloudflare, and Cisco, serve as popular alternatives to relying on local resolution, especially among more technically sophisticated users. We emphasize that the public DNS infrastructure offered by Google, Cloudflare, and Cisco all use IP anycast and are backed by highly distributed networks [57]. For example, DNS resolution requests to the fixed IP address of Google’s Public DNS resolvers will often be routed to a resolver located close to the requesting client [60].

Compared with local DNS resolution and with resolution via large, public DNS providers, resolution via SDNS resolvers causes the requests and responses to transit longer network paths. We confirm this by counting the number of autonomous systems (ASes) traversed between clients and (1) Google’s and Cloudflare’s public resolvers and (2) 108 identified SDNS resolvers. We determine the number of ASes by performing traceroutes and using the utility’s built-in IP-to-ASN mapping, and then counting the unique ASes observed in the reported network paths. More AS traversals indicate longer paths and consequently increased vulnerability to eavesdropping, since more organizations have the ability to observe the traffic. We place our traceroute clients in five continents, and report our results in Table 4.7. We find that the average number of AS traversals between the client and its chosen DNS
resolver increases when the client elects to use an SDNS resolver. The relative increase in the number of AS traversals ranges from 13% (Japan) to a near tripling in length (Belgium), relative to using Google’s or Cloudflare’s public DNS service; in the United States, the average number of ASes that observe the DNS requests increases by 55% when SDNS is used. Finally, we note that the use of distant DNS resolvers has been found to be a significant threat to privacy in the context of Tor [60]. We emphasize, however, that unlike with Tor, SDNS users send all DNS requests to potentially distant DNS resolvers, not just those that are produced when temporarily browsing anonymously with a specialized browser.

4.7.2 Eavesdropping on Proxies

SDNS customers are also exposed to an increased risk of eavesdropping through the use of the proxies themselves. Communicating via a proxy increases the surface area for eavesdropping. While directly accessing sites generally uses the shortest paths in terms of the number of autonomous systems traversed [47], relaying traffic through an SDNS proxy requires that it first be transmitted to the proxy and that the proxy separately transmit it to its destination. This process produces longer paths that are more vulnerable to eavesdropping.

These long paths are especially risky in the case of SDNS services since connections between users and their proxies are not encrypted. A user may use HTTPS to achieve end-to-end confidentiality of content with the visited website, but the widespread use of SNI allows an eavesdropper situated either between the user and the proxy or between the proxy and the destination to learn the hostnames of all requested URLs.

How much the eavesdropper can learn from this leakage mainly depends on what traffic the SDNS provider proxies. At the extreme, the HideMyIP SDNS service proxies all web requests, regardless of the requested webpage. Clearly, this leaks significant information to an on-path, passive eavesdropper and causes the SDNS provider to incur a very high bandwidth cost.
However, even in cases where SDNS providers take steps to limit unnecessary proxying, they likely still leak substantial information about their users’ Internet browsing habits. The SDNS provider ultimately decides which domain names it will route to its proxies and which it will allow its users to access directly. As noted in §4.3, SDNS services can limit unnecessary proxying by only proxying the content required to make their supported channels work—for example, just those web objects that consider the client’s location and enforce the geoblock restrictions. This can become problematic when a supported channel runs its geo-ip checks on a large CDN and references it using a ubiquitous domain name. In one such case, we noted that Netflix runs one of its geo-ip checks on an Akamai node (akamaihd.net). This effectively requires the SDNS provider to proxy all content to akamaihd.net including that which is not related to any supported channel. For example, the SmartDNSProxy SDNS service proxies some Akamai-hosted webobjects on the Honda motorcars website, despite Honda not being a supported channel.

This “over-proxying” allows an eavesdropper situated between the client and the proxy to learn not only about visits to supported channels, but also other sites that happen to use the same CDN nodes as those channels. We note that although such information may be encrypted, the now-ubiquitous use of SNI may leak information about requested hostnames.

**Unadvertised Proxying**  Given SDNS providers’ opportunity to limit costs by only proxying domain names needed to support their advertised channels, we expected SDNS providers to support only the channels that they advertise. However, we additionally identified several instances in which this was not the case. To discern instances of unadvertised proxying, we queried SDNS resolvers for domains from the Alexa website rankings list, and then compared the results to a ground-truth dataset we collected by issuing queries from a distributed collection of RIPE Atlas proxies as well as queries using Google’s, Cloudflare’s, and our local
institution’s DNS resolvers. A more detailed explanation of our methodology is presented in § 4.4.

We find that four SDNS providers (SmartDNSProxy, TrickByte, Uflix and VPNUK) omit supported domains from their published channel lists.

For the most part, the uncovered unadvertised channels correspond to pornographic websites. We posit that these sites are intentionally omitted from SDNS providers’ websites to avoid detracting from the providers’ perceived legitimacy or professionalism. As with over-proxying of domains, the failure of SDNS providers to announce the proxying of these channels poses privacy risks to their customers, since accessing these sites likely traverses longer Internet paths than would occur via direct connections. Beyond the longer paths incurred, all proxied sites leak the SNI hostname of websites visited over a TLS connection. As such, the aforementioned passive, on-path eavesdropper can learn that these users access pornographic sites, which sites they visit, and the frequency at which they do so. Moreover, the SDNS provider can change which domains it proxies at any time, and without warning. As a result, the eavesdropper could potentially gain insights into additional aspects of users’ browsing behavior.

4.8 Authentication and Authorization Failures

As discussed in §4.3, the workflow of SDNS is a two-step process: (i) upon receiving a DNS resolution request for a supported channel, an SDNS resolver returns the IP address of one of its proxies, and (ii) upon receiving HTTP/S requests from the end user, the proxy then either inspects the Host HTTP header or the TLS Server Name Indication (SNI) extension field to infer the destination, and then forwards TCP traffic to and from the location inferred. Critically, SDNS providers should perform authentication (is the requesting user a registered customer?) and authorization (should the traffic to the requested site be proxied?) at both of the above steps.

---

8We exclude HideMyIP from this analysis, since it proxies all connections regardless of destination.
In prior work, Fifield et al. used custom ZMap scans [38] to identify approximately 2500 open SNI proxies in the wild that used SNI introspection to proxy HTTPS traffic for arbitrary Internet users [43]. We compared our list of identified SDNS proxies to the open SNI proxies found by Fifield et al., and discovered four IP addresses on both lists, indicating that at least some proxies operated by SDNS services do not properly perform authentication. That is, they allow non-paying customers to directly use their proxies to relay traffic (e.g., to bypass geoblocking). We confirmed that the SDNS proxies allowed non-registered Internet users to proxy content by manually setting the SNI header in HTTPS requests originating from an IP that is not registered with the SDNS service. In all cases, the proxies retrieved the requested content.

To explore whether authentication failures were due to infrequent configuration errors on a small subset of a service’s proxies or endemic misconfigurations across all of its proxies, we identified additional proxy servers for eight SDNS providers (see Table 4.8). To find additional proxies, we noted that SDNS proxy servers sometimes presented distinctive error messages when accessed directly over HTTP, without a modified Host HTTP or SNI header that indicated an alternate destination. (Fifield et al. made a similar observation of open SNI proxies [43].)

These error messages often warned the user of a DNS misconfiguration and directed the user back to the SDNS provider’s website. We queried censys.io for this distinctive text to discover more potential proxies.

Table 4.8 shows the number of proxies we identified for each service, along with whether the proxies were restricted to SDNS customers (open circles) or functioned as open proxies (darkened circles). We tested whether a proxy was open by specifying Host HTTP headers for non-encrypted web traffic and TLS SNI headers for encrypted web traffic in an attempt to proxy. We found that CactusVPN, HideIPVPN, and SmartyDNS all had endemic authentication errors; all of their proxies functioned as open SNI proxies for any requesting Internet
oddly, we found no instances of open proxying for unencrypted traffic, even among those three providers. The authentication checks—which are based solely on the requestor’s IP address—are performed only for HTTP proxying. We note that, depending upon the computer platform, Google reports that between 74% to 94% of web requests using the Chrome browser are over HTTPS [56], suggesting that the failure to authenticate HTTPS requests is sufficient for non-customers to access the majority of the web.

We additionally checked whether the identified proxies would forward traffic to any domain, or limit proxying to its supported channels as determined by its responses to DNS resolution requests with a proxy’s IP address. We use the term universal to refer to proxies that forward traffic to any domain, and note that their presence in a given SDNS provider’s infrastructure indicates the provider’s failure to properly check the authorization of its proxying requests.

As shown in Table 4.8, we find that the identified proxies operated by CactusVPN, HideIPVPN, ibVPN, SmartyDNS, and VPNUK are all universal. All open proxies are also universal proxies, although the reverse does not hold for ibVPN and VPNUK. Proxies that are open and universal (CactusVPN, HideIPVPN, and SmartyDNS) allow any Internet user to proxy HTTPS traffic to any site, without having to register (or pay) for the SDNS service.

4.9 INFORMATION Leakage THROUGH DNS Probing

Most SDNS providers advertise a number of channels (i.e., web sites) for which they will proxy traffic. In this section, we describe how DNS cache probing techniques can be used to infer both (i) the relative popularity of channels among a service’s customers and (ii) the number of users of an SDNS service. Channel popularity allows us to gauge which sites’ geo-filters are most often bypassed, while the number of users of an SDNS service enables us to estimate the revenue and general profitability of the service. SDNS services do not publish statistics that describe which channels are actually accessed by their customers, nor do they provide the relative popularity of the channels that are accessed. The DNS probing
Table 4.8: SDNS Authorization Failures Leading to Open and Universal Proxys.

Occurrence of open and universal proxies, by SDNS provider, for both HTTP and SNI proxy methods. Uflix, Trickbyte, and SmartDNSProxy shared 10 proxy servers; SmartDNSProxy and Trickbyte shared an additional seven proxies; and CactusVPN and SmartyDNS used the same five proxies. Moreover, among the SDNS providers studied, we observed that, for each protocol supported, an SDNS provider’s proxies either were all open/universal, or none of them were.

<table>
<thead>
<tr>
<th>Provider</th>
<th>Confirmed Proxies</th>
<th>Open (HTTP/HTTPS)</th>
<th>Universal (SNI/HTTPS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CactusVPN</td>
<td>5</td>
<td>○</td>
<td>●●●●</td>
</tr>
<tr>
<td>HideIPVPN</td>
<td>3</td>
<td>○</td>
<td>●●●</td>
</tr>
<tr>
<td>IBVPN</td>
<td>1</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>SmartDNSProxy</td>
<td>36</td>
<td>○</td>
<td>○○○</td>
</tr>
<tr>
<td>SmartyDNS</td>
<td>5</td>
<td>○</td>
<td>●●●</td>
</tr>
<tr>
<td>Trickbyte</td>
<td>18</td>
<td>○</td>
<td>○○○</td>
</tr>
<tr>
<td>Uflix</td>
<td>14</td>
<td>○</td>
<td>○○○</td>
</tr>
<tr>
<td>VPNUK</td>
<td>17</td>
<td>○</td>
<td>●●○</td>
</tr>
</tbody>
</table>

○ none of the service’s tested proxies operated in this mode
● all of the service’s tested proxies operated in this mode

Techniques described in this section provide a first glimpse as to how SDNS customers use these services.

4.9.1 Inferring Channel Requests

To identify the channels requested by SDNS customers, we use the DNS cache snooping technique introduced by Grangeia [24, 58] to determine whether or not the zone record for a hostname is in the cache of a resolver.
By default, clients almost always set the Recursion Desired (RD) bit when sending queries to DNS servers. This is true of many major operating systems (including OSX, Windows, and Linux) and is intended to allow for the possibility of recursive DNS resolution. In brief, the RD bit indicates to the resolver that the client prefers that the resolver perform a recursive DNS lookup. Grangeia’s cache snooping technique leverages the behavior that when the RD bit is not set, DNS servers (i) respond with the resolved IP address if the entry is in its cache and (ii) return either an error or the root name servers for the requested domain if it is not. (We note that returning the root name servers is the expected behavior for iterative DNS resolutions.) By setting the RD bit to zero, we definitely learn whether the requested hostname is in the resolver’s cache.

Figure 4.3 shows the presence and absence of the hostnames for advertised channels on three SDNS providers over an approximately 5.25 day period beginning on August 21, 2019. We probed each SDNS provider’s cache once per hour during this period. (We performed the experiment for other SDNS providers and obtained similar results; they are omitted for brevity.) Specifically, we examined the cache of the first DNS resolver that was listed on the webpages of the three tested SDNS providers. We observe that (i) most of the domain names associated with the advertised channels never appeared in the resolvers’ caches and (ii) the few sites that did appear, did so consistently. For example, less than 24% (61 out of 256) of the sites supported by the SmartyDNS proxy ever appeared in its cache; TrickByte had the highest cache saturation with 61% (50 out of 82) of its supported channels appearing at least once in its cache during our measurement period. In summary, while SDNS providers advertise support for a large number of channels, our findings suggest that customers regularly use only a modest fraction of those offerings.

4.9.2 Deriving Channel Popularity

Based on our prior analysis, we sought to understand the relative popularity of channels provided by SDNS. We estimate how often an SDNS provider’s customers request each of
the provider’s supported channels by assuming the rate at which the SDNS provider resolves requests for a particular hostname is indicative of frequency at which its customers request it.

To perform this analysis, we use the popularity inference technique of Rajab et al. [100], which operates as follows: a request for resolving hostname $H$ is sent to a DNS resolver $D$. In its reply, $D$ returns the time ($TTL_d$) at which the entry for $H$ will be expunged from its cache. The maximum possible TTL ($TTL_{max}$) can be obtained by querying the authoritative name server for $H$. Rajab et al.’s technique issues a probe for $H$ once per $TTL_{max}$, which allows for computing the refresh time (the time at which the cache entry for $H$ was most recently refreshed) as $T_r = T_p - (TTL_{max} - T_i)$ where $T_p$ is the time of the probe. This allows for the computation of the average rate $\lambda$ at which $H$ is requested from $D$ as

$$\lambda \approx \frac{R}{\sum_{i=1}^{R}(T_{r_i} - T_{r_{i-1}} - TTL)}$$

where $R$ is the number of probes for $H$ [100].
We implemented Rajab et al.’s algorithm and deployed it on 11 resolvers belonging to 11 different SDNS providers (i.e., we used one resolver per SDNS provider). Three of the 11 resolvers reported erratic or otherwise erroneous TTLs, and we excluded these resolvers from our analysis. For the remaining eight SDNS services, the average aggregate request rates, measured in requests per hour, are listed in Table 4.9. We report the 10 most frequently requested hostnames for each SDNS service. We additionally compute 95% confidence intervals of $\lambda$ by applying the central limit theorem [100]; this allows us to gauge our confidence in the results based, in part, on the number of probes we performed. (Hostnames that have a large $\text{TTL}_{\text{max}}$ value resulted in fewer probes.)

Our findings reveal that streaming video—the target offering for many of the SDNS providers—is by far the most common destination for SDNS customers. Interestingly, SDNS providers commonly resolve queries for popular news (cnn.com) and social media (instagram.com) sites. This suggests that SDNS customers regularly use SDNS resolvers and do not reserve their use for accessing geoblocked content.

In §4.10, we use similar techniques to estimate the number of customers and revenue for several SDNS providers.

**Limitations to Popularity Measures** To minimize the volume of our requests, we target only one DNS resolver for each SDNS provider. Probing the unexamined DNS resolvers may yield different results. Additionally, as shown in Table 4.9, the confidence intervals can be large, sometimes overwhelming a site’s estimated arrival rate ($\lambda$). In such cases, these results should be viewed with some skepticism. Finally, our results assume DNS servers correctly follow the DNS protocol as described in the RFC [91]. Although, by definition, SDNS resolvers do not always return correct DNS responses, we observe that the tested SDNS resolvers appear to generally adhere to the DNS RFC, with the sole exception of returning false IP addresses for proxied channels.
4.10 Estimating the Number of Customers and Revenue

Again borrowing from the technique of Rajab et al. [100], we can use the average request rate ($\lambda$) to form a rough approximation of the number of users ($n$) of an SDNS provider. We denote $\lambda$(site) as the value of the aggregated (total) average request rate for a given site. Further, let $\lambda_c$(site) be the expected request rate for a client accessing the site. That is, while $\lambda$(site) denotes the total requests per unit time for the site, $\lambda_c$(site) is the number of requests due to a given user. Then, assuming Gamma distributed arrival times, Rajab et al. shows that the number of users $n$ of an SDNS provider is:

$$n = \frac{\lambda(\text{site})}{\lambda_c(\text{site})}$$

Rajab et al. reports that $\lambda_c$(google.com) is 2.63 requests per hour [100]. We can compute $\lambda$(google.com) for the various SDNS providers using the technique described in the previous section, and thus compute $n$. Here, we perform our probes (i.e., DNS requests for google.com) over an approximately 11 hour period beginning on September 6, 2019. To limit the load on the SDNS providers’ resolvers, we send probes to a single resolver per SDNS provider.

Table 4.10 lists the empirically measured average request rate for google.com and the derived number of users for six SDNS services’ resolvers. (The remaining providers did not consistently respond properly to DNS requests to resolve google.com, and are excluded from the Table.) Note that the number of estimated users in Table 4.10 is based on traffic to a single resolver per service and thus likely undercounts the total number of users of a service.

Of the successfully tested providers, we find that CactusVPN has approximately 16K users using a single one of its resolvers, while the other SDNS services have significantly fewer users accessing their tested resolvers.

Using the pricing information presented in Table 4.1, we can then estimate the revenue for each SDNS provider by multiplying the estimated number of users by the price-per-user. This should be considered a conservative (low) estimate of the provider’s revenue, since our probing DNS requests target only the first listed DNS resolver for each SDNS provider.
We can estimate profit margins for an SDNS provider based on the expected costs of running proxy servers. Proxies relay content to/from supported channels, and we consider a near-worst case scenario in which all SDNS users continuously stream high-quality video. Here, we use Netflix’s reported bandwidth requirements of 3 GB/hour (6.67 Mbps) to estimate SDNS providers’ bandwidth needs. SDNS providers can easily support such rates with VPS providers. In particular, there are a number of VPS providers that provide uncapped (sometimes called *unmetered*) 1 Gbps links [52] for approximately $10 per month. We note that a single 1 Gbps link can support 150 SDNS customers who each consume 6.67 Mbps. The revenue from 150 SDNS customers far exceeds the bandwidth costs. For example, CactusVPN charges $4.99 per customer, per month, for a revenue of $748.50 and profit of $738.50 (after subtracting the $10 VPS cost) per 150 customers. The profit per customer is thus $4.92 per month, yielding a profit margin as high as 98.6%. Using these general assumptions, we provide estimates of the profits of SDNS providers in Table 4.10.

4.10.1 Limitations to Profit and Revenue Estimation

Our analysis relies on a number of assumptions, including the expected distributed arrival times and the accuracy of $\lambda_c$(google.com) reported by Rajab et al. in 2010. We note that the client enumeration attack presented in §4.6 constitutes a far more accurate method of determining the precise number of SDNS customers, although the attack only works for a subset of SDNS providers. (Due to obvious ethical concerns, we did not perform the enumeration attack described in §4.6 to measure SDNS usage.)

Additionally, as discussed above, a more complete revenue exploration of an SDNS service would include the infrastructure costs of resolvers, as well as the fact that not all proxies are fully utilized. Further, our analysis ignores the potentially high costs of customer support and maintaining an infrastructure for billing.
Many of the vulnerabilities identified in this chapter are inherent to the design of SDNS systems, and are difficult to remedy. In particular, the real-time SDNS user identification attack (§4.6.2)—which can also identify SDNS proxy servers—is effective because SDNS can only proxy traffic that first requires a DNS resolution. Fundamentally, the attack exploits the defining characteristic of SDNS services (i.e., the assignment of proxies via DNS resolution), and thus we believe it is unlikely that an SDNS provider can counter this inherent weakness.

In contrast, other vulnerabilities we identify can be remedied by purveyors of vulnerable SDNS systems. The client enumeration attack (§4.6.1) exploits an implementation flaw in some SDNS systems, and can be Remedied; in fact, one provider implemented a fix after we disclosed the attack.

The increased risk of traffic analysis (§4.7) is similar to the risk that arises from using VPN services: longer Internet paths generally provide more opportunities for eavesdropping. However, unlike with VPNS, SDNS has no easy on/off switch, and we suspect that many SDNS users configure their computers to use SDNS resolvers and do not restore their settings after using SDNS. This “set and forget” functionality is unusual for VPNS, which typically require user interaction to enable. We conjecture that SDNS use provides a more persistent level of susceptibility to eavesdropping than VPNS due to the likely longevity of using SDNS resolvers.

Encrypting DNS traffic between a client and the SDNS resolver using either DoH [63] or DoT [65] mitigates some of the eavesdropping risks. However, encrypted DNS is still subject to traffic analysis which can leak the sites being resolved [25, 64, 110] to on-path eavesdroppers. Perhaps more importantly, not all major operating systems support DoH or DoT. We believe at least for the short-term that it is unlikely that SDNS providers would add support for DoH/DoT, since doing so may increase the level of technical sophistication required to configure SDNS.
The authentication and authorization failures (§4.8) leverage poor design decisions by the vulnerable SDNS services. Applying IP-based authentication at both the resolver and the proxy prevent unauthorized use. (However, it is worth emphasizing that IP-based authentication does not provide strong authentication.)

Finally, the exposure to analytics (§4.9) uses DNS cache probing techniques that are generally applicable to DNS resolvers. They are arguably especially problematic however in the SDNS setting since the sole purpose of SDNS services is to bypass geoblocking, and hence determining how often these services are used for each channel could be useful to assess the potential criminal culpability or legal liability of the providers. To prevent such information leakage, DNS resolvers whether SDNS resolvers or ordinary resolvers could advertise the maximum TTL value, although such behavior would clearly violate the DNS specification [91].

4.12 Ethical Considerations

We sought to minimize risk, both to SDNS users and to the services themselves. Our experiments were guided by the principles outlined in the Menlo Report [35]:

4.12.1 Respect for Persons

We avoided causing harm to users through our measurements and proof-of-concept attacks by not targeting specific individuals. In validating the client enumeration attack in §4.6.1, we used a small-scale proof-of-concept in which we spoofed only our own IP addresses. We did not attempt to discern whether any IP addresses, other than those operated by the authors, belonged to customers of the vulnerable SDNS services. We did not issue more than a handful of queries to the SDNS provider’s resolver, and our DNS queries conformed to the DNS standard [91].

When identifying SDNS usage rates and the popularity of channels (§4.9), our measurements consisted of sending a relatively low volume (one request per hostname per the
hostname’s TTL_{\text{max}}) of well-formed DNS queries to a resolver. This volume of requests is negligible compared to the request arrival rate of SDNS providers (see Table 4.9). This measurement provides respect for individuals and persons in that it does not directly interfere with the normal activities of the SDNS service. We used SDNS resolvers exactly as they were intended to function, and merely inspected the returned result (i.e., its TTL) to derive statistical inferences.

4.12.2 Beneficence

At all stages of our study, we sought to reduce harm and maximize the benefits of the research. Foremost, we designed our experiments to avoid overwhelming the public DNS or SDNS resolvers by rate limiting our requests, and we only submitted well formed DNS, HTTP, and HTTPS requests throughout.

Most relevant, after identifying the client enumeration attack, we performed responsible disclosure and notified the operators of the two affected services (VPNUK and ibVPN). We received a response from ibVPN the following day, informing us that they had partially mitigated the attack (see §4.6.1). The goal of disclosing the attack was to decrease the associated risks identified and increase the overall benefits of the research.

4.12.3 Justice

Our measurements are just in that we spread out all probes to a broad set of SDNS providers, treating each equally in our experimentation without targeting specific services over others.

4.12.4 Respect for Law and Public Interest

We designed our study to be both lawful and in the public interest. We paid or used free-trials for all of the measured SDNS providers; we did not use these services surreptitiously. Additionally, we considered and reported on the privacy and security ramifications of using these services, which is in the public interest, and when we identified potential vulnerabilities, we responsibly disclosed them to the SDNS providers.
4.13 Summary

This chapter presents one of the first studies of smart DNS (SDNS) services in the wild. As a tool whose sole intended use is geoblock evasion, and that is customized to unblock with minimal overhead costs, its architecture, functionality and security and privacy properties provide unique insights into both the shortfalls of geoblocking setups commonly used by websites and their underlying IP-geolocation functionalities, as well as those common amongst unblockers more generally. In particular, we show that:

1. Popular websites that geoblock sometimes rely on their CDN providers for their geofiltering functionality. Through our observations of overproxying amongst SDNS providers (§4.7.2), we find additional evidence indicating that CDNs likely play a role in the ongoing proliferation of geoblocking and possibly of IP-geolocation more generally throughout the Internet, particularly amongst online streaming services. In at least some cases, streaming content providers that rely on their CDNs for this functionality, embed queries to domain names controlled by their CDNs for the underlying IP-geolocation information they need to geoblock. The appearance of queries to these domain names also appear within the embedded resources of other, seemingly unrelated websites, raises the question of whether these other websites also use IP-geolocation to customize their behavior (be it for language settings, price listings etc.). Assuming this is the case, SDNS usage could unintentionally hinder web servers’ ability to customize their behavior to SDNS users’ respective locations.\(^9\)

2. Website owners can often detect when an unblocker is being used to circumvent their geofiltering. As we show in §4.5, content providers can trivially detect SDNS users who access their sites, identify these users’ IP addresses and enumerate all customers who are registered for these SDNS services regardless of whether or not they access the

\(^9\)It is worth noting that we are not currently aware of the pervasiveness of this type of customization across different types of websites, nor can we infer how use of a specific SDNS service will impact different websites’ performances. For this reason we believe overproxying and its effects warrant additional study.
content provider’s website. This can have a number of different consequences including website’s cracking down on unblocker usage thus diminishing these tools’ overall efficacy at mildest, and at worst, a breach of SDNS users’ privacy.

3. Commercial unblockers can be especially prone to misconfigurations. In some cases this can pose a threat to their own profitability. We specifically identify a number of authentication and authorization errors in the setup of SDNS, with many services relying only on IP-based authentication at the DNS server and neglecting to perform these checks at the proxy server. The failure to properly authenticate and authorize users effectively transforms SDNS providers’ proxies into a distributed network of open SNI proxy servers. We describe a straightforward method of discovering these open proxies and discuss how unscrupulous users could use SDNS services while bypassing payment.

4. In some cases these misconfigurations can pose a threat to users’ online security and privacy. In the case of SDNS systems, the deproxying attack identified in §4.5 is inherent to these systems’ construction. However, misconfigurations such as the one described in §4.6.1 further exacerbate the threat posed by these architectural flaws. Specifically, some settings of SDNS allow any arbitrary third party to enumerate all customers of an SDNS service, even when those users are offline. These vulnerabilities were confirmed and repaired by an SDNS provider after responsible disclosure.

As such, our findings reaffirm that the increased adoption of geoblocking poses security and privacy risks to blocked Internet users who seek to evade it using unblockers like SDNS.
Table 4.9: SDNS Most Popular Channels (Derived). Derived most popular channels and average estimated resolution rate (\( \lambda \)), in requests per hour.

<table>
<thead>
<tr>
<th>Site</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tvland.com</td>
<td>47526 ±65</td>
</tr>
<tr>
<td>player.pl</td>
<td>47462 ±131</td>
</tr>
<tr>
<td>hbogo.com</td>
<td>47413 ±76</td>
</tr>
<tr>
<td>pandora.com</td>
<td>47339 ±93</td>
</tr>
<tr>
<td>comedycentral.com</td>
<td>47339 ±158</td>
</tr>
<tr>
<td>sonycrackle.com</td>
<td>39475 ±2385</td>
</tr>
<tr>
<td>theloop.ca</td>
<td>39099 ±4992</td>
</tr>
<tr>
<td>absoluteradio.co.uk</td>
<td>38714 ±217K</td>
</tr>
<tr>
<td>amazon.co.uk</td>
<td>36712 ±2224</td>
</tr>
<tr>
<td>bleacherreport.com</td>
<td>30703 ±6217</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>amazon.com</td>
<td>95694 ±401509</td>
</tr>
<tr>
<td>youtube.com</td>
<td>79222 ±79257</td>
</tr>
<tr>
<td>foxsportsgo.com</td>
<td>70913 ±5113</td>
</tr>
<tr>
<td>bloomberg.com</td>
<td>6726 ±5174</td>
</tr>
<tr>
<td>disneylife.com</td>
<td>6498 ±673</td>
</tr>
<tr>
<td>funimation.com</td>
<td>6434 ±247</td>
</tr>
<tr>
<td>theloop.ca</td>
<td>46122 ±278</td>
</tr>
<tr>
<td>travelchannel.com</td>
<td>46036 ±874</td>
</tr>
<tr>
<td>amcvtv.com</td>
<td>4587 ±44</td>
</tr>
<tr>
<td>vinaplay.se</td>
<td>45948 ±59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CactusVPN</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>foxsports.com.au</td>
<td>1.7M ±2.7M</td>
</tr>
<tr>
<td>zdf.de</td>
<td>395633 ±24824</td>
</tr>
<tr>
<td>m6replay.fr</td>
<td>393483 ±2741</td>
</tr>
<tr>
<td>tsn.ca</td>
<td>91665 ±1662</td>
</tr>
<tr>
<td>rds.ca</td>
<td>89007 ±1163</td>
</tr>
<tr>
<td>tennischannel.com</td>
<td>72537 ±2066</td>
</tr>
<tr>
<td>libonow.com</td>
<td>66332 ±14379</td>
</tr>
<tr>
<td>itv.co</td>
<td>59623 ±15541</td>
</tr>
<tr>
<td>itv.com</td>
<td>44405 ±872</td>
</tr>
<tr>
<td>foxsoccer2go.com</td>
<td>34031 ±209</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IB VPN</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>starzplay.com</td>
<td>90721 ±43758</td>
</tr>
<tr>
<td>cartoonnetwork.com</td>
<td>88345 ±216417</td>
</tr>
<tr>
<td>alctv.com</td>
<td>87648 ±91137</td>
</tr>
<tr>
<td>ewtv.com</td>
<td>79143 ±74440</td>
</tr>
<tr>
<td>ewseed.com</td>
<td>78446 ±51686</td>
</tr>
<tr>
<td>indieflix.com</td>
<td>71231 ±219415</td>
</tr>
<tr>
<td>theonion.com</td>
<td>35655 ±11610</td>
</tr>
<tr>
<td>teamcoco.com</td>
<td>35201 ±137079</td>
</tr>
<tr>
<td>tlc.com</td>
<td>35200 ±1.52M</td>
</tr>
<tr>
<td>showtime.com</td>
<td>35196 ±118214</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Keenow</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>player.pl</td>
<td>423944 ±829</td>
</tr>
<tr>
<td>hbogo.com</td>
<td>421225 ±683</td>
</tr>
<tr>
<td>pandora.com</td>
<td>413256 ±1182</td>
</tr>
<tr>
<td>tvland.com</td>
<td>406346 ±4193</td>
</tr>
<tr>
<td>comedycentral.com</td>
<td>404203 ±5509</td>
</tr>
<tr>
<td>amazon.co.uk</td>
<td>306602 ±209K</td>
</tr>
<tr>
<td>cnbc.com</td>
<td>252442 ±93K</td>
</tr>
<tr>
<td>sling.com</td>
<td>208366 ±34462</td>
</tr>
<tr>
<td>nickjr.com</td>
<td>202391 ±46491</td>
</tr>
<tr>
<td>foxsports.com</td>
<td>197627 ±20198</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SmartyDNS</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>docclub.com</td>
<td>79823 ±67529</td>
</tr>
<tr>
<td>discovery.com</td>
<td>78740 ±48817</td>
</tr>
<tr>
<td>showcase.ca</td>
<td>74609 ±58135</td>
</tr>
<tr>
<td>tvplayer.com</td>
<td>66103 ±90796</td>
</tr>
<tr>
<td>zattoo.com</td>
<td>59129 ±54614</td>
</tr>
<tr>
<td>syfy.com</td>
<td>34851 ±31357</td>
</tr>
<tr>
<td>nbc.com</td>
<td>33614 ±6276</td>
</tr>
<tr>
<td>history.com</td>
<td>33599 ±2476</td>
</tr>
<tr>
<td>rdio.com</td>
<td>33331 ±3398</td>
</tr>
<tr>
<td>klowditv.com</td>
<td>33275 ±67988</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SmartDNSProxy</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>amazon.co.uk</td>
<td>754402 ±108K</td>
</tr>
<tr>
<td>oxygen.com</td>
<td>497202 ±345K</td>
</tr>
<tr>
<td>magine.com</td>
<td>341406 ±468</td>
</tr>
<tr>
<td>songza.com</td>
<td>324251 ±4487</td>
</tr>
<tr>
<td>vtele.ca</td>
<td>318099 ±4708</td>
</tr>
<tr>
<td>abema.tv</td>
<td>285243 ±3549</td>
</tr>
<tr>
<td>cbc.ca</td>
<td>285113 ±307K</td>
</tr>
<tr>
<td>telemundo.com</td>
<td>271029 ±211K</td>
</tr>
<tr>
<td>rds.ca</td>
<td>268518 ±16790</td>
</tr>
<tr>
<td>tennischannel.com</td>
<td>261632 ±15117</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IronSocket</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>instagram.com</td>
<td>2708804 ±97K</td>
</tr>
<tr>
<td>epix.com</td>
<td>116632 ±388</td>
</tr>
<tr>
<td>vh1.com</td>
<td>116564 ±7429</td>
</tr>
<tr>
<td>discovery.com</td>
<td>113888 ±21888</td>
</tr>
<tr>
<td>cnbc.com</td>
<td>113722 ±9419</td>
</tr>
<tr>
<td>history.com</td>
<td>111340 ±2690</td>
</tr>
<tr>
<td>amc.com</td>
<td>109885 ±2764</td>
</tr>
<tr>
<td>aotv.com</td>
<td>108331 ±1187</td>
</tr>
<tr>
<td>beinsports.com</td>
<td>100153 ±1044</td>
</tr>
<tr>
<td>cnn.com</td>
<td>101036 ±187K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DNSTrick</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>docclub.com</td>
<td>79823 ±67529</td>
</tr>
<tr>
<td>discovery.com</td>
<td>78740 ±48817</td>
</tr>
<tr>
<td>showcase.ca</td>
<td>74609 ±58135</td>
</tr>
<tr>
<td>tvplayer.com</td>
<td>66103 ±90796</td>
</tr>
<tr>
<td>zattoo.com</td>
<td>59129 ±54614</td>
</tr>
<tr>
<td>syfy.com</td>
<td>34851 ±31357</td>
</tr>
<tr>
<td>nbc.com</td>
<td>33614 ±6276</td>
</tr>
<tr>
<td>history.com</td>
<td>33599 ±2476</td>
</tr>
<tr>
<td>rdio.com</td>
<td>33331 ±3398</td>
</tr>
<tr>
<td>klowditv.com</td>
<td>33275 ±67988</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TrickByte</th>
<th>( \lambda ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CactusVPN</td>
<td></td>
</tr>
<tr>
<td>IB VPN</td>
<td></td>
</tr>
<tr>
<td>Keenow</td>
<td></td>
</tr>
<tr>
<td>SmartyDNS</td>
<td></td>
</tr>
<tr>
<td>SmartDNSProxy</td>
<td></td>
</tr>
<tr>
<td>IronSocket</td>
<td></td>
</tr>
<tr>
<td>DNSTrick</td>
<td></td>
</tr>
<tr>
<td>TrickByte</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.10: Estimated SDNS Resolver Usage and Provider Monthly Profit. The estimated number of users of a single SDNS resolver for various SDNS providers, and the estimated monthly profit.

<table>
<thead>
<tr>
<th>Service</th>
<th>Rate (λ)</th>
<th>Est. Users (n)</th>
<th>Est. Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>CactusVPN</td>
<td>41,119</td>
<td>15,635</td>
<td>$76,977</td>
</tr>
<tr>
<td>DNStrick</td>
<td>1,794</td>
<td>682</td>
<td>$3,330</td>
</tr>
<tr>
<td>HideIP VPN</td>
<td>2,127</td>
<td>809</td>
<td>$3,952</td>
</tr>
<tr>
<td>SmartyDNS</td>
<td>6,389</td>
<td>2,429</td>
<td>$11,741</td>
</tr>
<tr>
<td>TrickByte</td>
<td>8,269</td>
<td>3,144</td>
<td>$9,190</td>
</tr>
<tr>
<td>Unlocator</td>
<td>3,565</td>
<td>1,356</td>
<td>$6,622</td>
</tr>
</tbody>
</table>
Chapter 5

Geoblocking and its Evasion as Perceived by Users of Popular Unblockers

Given how common it is for people to want to evade geoblocking and the threats unblockers, such as SDNS can pose, it is unclear to what extent users of these systems are cognizant of the associated security and privacy costs. In our initial measurement study of SDNS systems we observe a disconnect between the risks we uncover, and how these services were presented to would be users. In particular, we note that, similar to Khan et al.'s observation about VPN providers [75], many SDNS providers' websites appear to advertise more security than these services actually offer. When coupled with the previous chapter's work indicating SDNS's widespread usage and popularity, SDNS systems appear to pose a significant privacy and security threat to geoblocked Internet users. This inspired our work in this chapter, which seeks to understand users' perceptions of their geoblock evasion service providers, and to ascertain the extent to which they believe using these unblockers offers them additional security or privacy. Since SDNS specifically caters to geoblock evasion, its users are an ideal sample for studying user's perceptions of these types of tools. To that end, this chapter focuses on SDNS users' perceptions of these services. In so doing, it explores four main research questions: (1) what motivates users to use SDNS?; (2) how do users select which

This chapter is based on the paper
User Perceptions of the Privacy and Usability of Smart DNS, which appeared at the 2022 Annual Computer Security Applications Conference (ACSAC 2022). The paper’s complete citation is given below:

SDNS provider to use?; (3) to what degree do SDNS users trust these services?; and (4) how well do SDNS users actually understand SDNS functionality?

To answer these questions, we conducted an online survey of 63 SDNS users, recruiting participants from topical subreddits (Reddit discussion boards) and Prolific. Participants were asked about their histories and experiences with SDNS services, their understanding as to how these systems operated, and their perceptions of SDNS’ trustworthiness and ethics.

We find that most of our study participants had misconceptions about how SDNS services function. Alarmingly, many conflated SDNS with VPNs, and considered the former to be a privacy-enhancing technology that provided additional layers of encryption and/or sender anonymity—in actuality, SDNS services do neither. Many were prone to put undue trust in the protections that SDNS services purport to offer.

Moreover, very few of the SDNS users who took our study considered the privacy and security risks of using SDNS. While many participants incorrectly understood that SDNS services would bolster their privacy, very few participants considered that using SDNS could diminish their security and privacy.

Interestingly, most participants viewed their use of SDNS services to bypass geofiltering as ethical. Several participants justified their use of SDNS by noting that the internet should be open and free from geography-based discrimination, and SDNS was a technology that helped achieve this ideal. Most participants similarly believed that using SDNS was legal, although a large portion of participants were uncertain of its legality.

The main findings of our study—that participants often did not understand how SDNS services functioned, and that they overestimated its privacy protections while underestimating the risk that using these services posed—should be construed as a strong signal that these services deserve more study. Our survey suggests that SDNS users may be unknowingly risking their security and privacy, and that much more user education is needed.
5.1 Methodology

To ascertain users’ understandings and perceptions of SDNS services, we conducted an online study. In this section, we describe the study’s recruitment and screening procedures, as well as its amended eligibility criteria, our ethical considerations when conducting this study, and the study’s limitations. For completeness, all components of the survey are included in Appendix A.

5.1.1 Study Procedure

The main components of our study included a pre-screening phase, and a main survey:

Pre-screening Participants  To ensure familiarity with SDNS, we asked participants a series of screening questions. Respondents whose answers indicated they were not familiar with SDNS were not allowed to complete the main survey.

We initially required respondents to either have experience using SDNS, as a current or previous SDNS user, or to be seriously considering using the service within one month of completing the survey\(^1\). As described in more detail in §5.1.2, participants were recruited across two separate platforms (Reddit and Prolific), which necessitated two distinct presentations of screening questions (S1 through S8)\(^2\). Despite this, both groups of participants were ultimately asked the same questions and were required to meet the same requirements to be eligible to complete the main survey.

Main Survey  The main survey consisted of questions covering the following topics:

1. SDNS impact on online security and privacy: To ascertain how participants believed that SDNS services affected their security and privacy, we began by asking participants whether they thought the service made their internet browsing more or less secure,

\(^1\)Responses from participants who did not have experience using SDNS were later excluded from analysis. We explain our reasoning for this decision in §5.1.2

\(^2\)Survey questions are referenced in bold typeface and can be found in Appendix A.
and whether it provided additional protections or posed a risk to their online security and/or privacy (questions M1-M4).

2. **SDNS functionality:** Participants were then asked a set of more specific questions about how SDNS systems operate. These questions were aimed at getting more detailed insights into participants’ understandings and misconceptions. Participants were presented with two sets of Likert scale questions that asked whether they agreed or disagreed with statements describing SDNS’s proxying behavior (M5) and ability to conceal customers’ IP addresses from websites (M6).

3. **Choice to use SDNS:** To gain context about participants’ SDNS usage, we then inquired about their goals in using SDNS. We surveyed participants about their motivations to specifically use this service, and whether they thought it was worth the effort required to set it up (M7-M11).

4. **Participant setup and usage of SDNS:** Next, respondents were asked about their setup and usage of SDNS services. These questions covered (1) whether their SDNS providers offered services in addition to SDNS (e.g., VPNs) and the extent to which participants used them (M12-M15); and (2) the types of devices on which participants set up SDNS and the extent to which they had SDNS enabled when browsing the web (M16-M18).

5. **Trustworthiness of SDNS providers:** Participants were asked whether they thought their chosen SDNS provider was trustworthy in general, and the steps they trusted their SDNS provider to take to ensure that (1) the service functioned as advertised and (2) that users’ security and privacy were safeguarded (M19-M21).

6. **Success evading detection:** Given the ongoing cat-and-mouse dynamic between content providers and SDNS services, participants were then asked about their experience accessing geoblocked content using these services. These questions focused on whether participants had been blocked by a content provider due to their use of SDNS, how
often they were caught, and how easy they thought it was for a content provider to
detect that they were using SDNS (M22-M26).

7. Ethics and legality of using SDNS: Finally, participants were asked whether they
thought using SDNS to access geoblocked content was ethical and/or legal, and to
explain their opinions (M27-M28).

8. Demographics (Reddit participants only): Participants who were recruited through
Reddit were then asked demographic questions (D1-D6).

Analysis of Responses  As we explain in more detail in §5.1.2 and §5.1.4, five participants’
responses had to be removed from quantitative analysis, and one response had to be removed
from qualitative analysis despite our initial screening for participant eligibility. Given our
small sample size (n = 63 for qualitative analysis, n = 58 participants in both qualitative and
quantitative analysis), we perform a qualitative analysis of survey responses. Qualitative,
or free-text responses were open-coded by a primary coder and were then evaluated by a
secondary coder (Cohen’s $\kappa \geq 0.76$). Our code book is included as Appendix B.

5.1.2 Recruitment and Eligibility Criteria

As noted in §5.1.1, participants were recruited across two groups: One group was recruited
on Reddit through posts that advertised our study on r/samplesize³, r/SurveyExchange⁴
and r/TakeMySurvey⁵ between February 1, 2021 and March 25, 2021. Upon completion of
the survey, Reddit participants were given the opportunity to enter a raffle to win a $50
USD Amazon gift card with minimum odds of winning of 1/20.

To determine eligibility, Reddit participants were screened using conditional navigation.
That is, they were presented with a single survey that began with screening questions S1
through S8. Before completing the survey, Reddit respondents were warned that the survey
included screening questions and attention checks, and, if determined to be unqualified,
they would be screened out of the survey and ineligible for remuneration. As such, Reddit respondents whose answers to screening question indicated that they neither had used SDNS as a current or previous user (S3), nor that they were not seriously considering using it within the next month (S4), were screened out mid-survey. Additionally, responses from Reddit participants that indicated familiarity with SDNS, but failed to identify the service’s main use—i.e., bypassing access restrictions to domains that performed geoblocking—were manually removed from consideration in real-time. As outlined in their informed consent, Reddit respondents who were screened out or submitted these removed entries were ineligible for remuneration.

The second group of participants was recruited using Prolific. Since Prolific does not allow participants to be screened out mid-survey, these respondents were given a screening survey in which they were asked these same screening questions, as well as demographic questions D1 through D6. Although included in the Prolific group’s screening survey, we note that respondent answers to the demographic questions were not considered when determining their eligibility to participate in this study. Those who were qualified, as we determined from their answers to the screening questions, were invited to return and take the main survey, which we described in more detail in §5.1.1 and whose complete contents can be found in Appendix A. Prolific participants were compensated $0.75 USD for completing the prescreen questionnaire and earned an additional $4.25 USD for completing the main survey. On average, Prolific participants took about 5 minutes to complete the prescreen, and 10 to 15 minutes to complete the main survey. Similarly, Reddit participants completed the combined survey in roughly 15 to 20 minutes.

Amended Eligibility Requirements Using the initial eligibility criteria described above, we surveyed 72 respondents across both groups (Reddit: n = 18; Prolific: n = 54). However, upon further scrutiny of responses, we found that responses from participants who did not have experience using SDNS either from current or previous usage of these

---

6https://www.prolific.co/
services, were often of lower quality. As such, once all participants had been compensated, these responses were excluded from analysis. Additionally, we found five responses that could not be included in our quantitative analysis; these consisted of four participants who indicated having used a service provider that provided VPN services but we discovered did not actually offer SDNS, and one participant whose answers reflected VPN usage but not that of SDNS. As we explain in more detail in §5.2 and §5.3, participant confusion over whether they had in fact used SDNS and not solely used a VPN was likely due to their general confusion about the differences between the two services. Since it is not uncommon for SDNS providers to also offer VPN services, and for VPNs to also be used to bypass geoblocks, we only removed these responses from quantitative analysis.

In total, 63 responses (Reddit: \( n = 11 \); Prolific: \( n = 52 \)) were included in qualitative analysis, and, of those 58 responses (Reddit: \( n = 10 \); Prolific: \( n = 48 \)) were also included in quantitative analysis. The final groups of participants had the demographic makeup described in Table 5.1.

5.1.3 Ethical Considerations

This study protocol was approved by the Georgetown University and The George Washington University Institutional Review Boards (IRBs) with approval numbers STUDY00002643 and NCR202561 respectively. All data collected was deidentified such that all participants’ personally identifiable information (PII) was removed, and responses were associated with random identifiers.

Since we ask participants whether they believe using SDNS to bypass geoblocking is ethical and/or legal, we also consider whether their responses to these two questions could constitute a confession of wrongdoing. This is not straightforward since determining the legality of using SDNS to bypass geoblocking appears to be quite complicated [114, 115].\(^7\)

\(^7\)Based solely on participant responses, we observe a general lack of clarity on the legality of using SDNS to bypass geoblocking, and that its legal status appears to vary across different regions’ regulations. However, we reiterate the caveat that we are neither lawyers, nor legal experts on this matter.
Table 5.1: Study Participant Demographics. Participant demographics for Reddit and Prolific participants included in both the quantitative and qualitative analyses. Prolific demographics exclude tallies from respondents who only completed the pre-survey.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Reddit sample</th>
<th>Prolific sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Participants</td>
<td>10</td>
<td>48</td>
</tr>
<tr>
<td>Gender: Male</td>
<td>7 (70%)</td>
<td>33 (69%)</td>
</tr>
<tr>
<td>Gender: Female</td>
<td>3 (30%)</td>
<td>14 (29%)</td>
</tr>
<tr>
<td>Gender: Prefer not to disclose</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Age: 18-24</td>
<td>3 (30%)</td>
<td>30 (63%)</td>
</tr>
<tr>
<td>Age: 25-34</td>
<td>4 (40%)</td>
<td>12 (25%)</td>
</tr>
<tr>
<td>Age: 35-44</td>
<td>2 (20%)</td>
<td>4 (8%)</td>
</tr>
<tr>
<td>Age: 45-54</td>
<td>0</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>Age: 65-74</td>
<td>1 (10%)</td>
<td>0 (%)</td>
</tr>
<tr>
<td>Less than high school degree</td>
<td>0</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>High school graduate, diploma, or equivalent</td>
<td>1 (10%)</td>
<td>15 (31%)</td>
</tr>
<tr>
<td>Trade/technical/vocational training/Associates’ Degree</td>
<td>1 (10%)</td>
<td>4 (8%)</td>
</tr>
<tr>
<td>Some college credit, no degree</td>
<td>2 (20%)</td>
<td>10 (21%)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>3 (30%)</td>
<td>12 (25%)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>3 (30%)</td>
<td>4 (8%)</td>
</tr>
<tr>
<td>Professional degree (e.g., J.D., M.D.)</td>
<td>0</td>
<td>1 (1.3%)</td>
</tr>
<tr>
<td>Doctoral Degree:</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Annual income: Less than $10,000</td>
<td>2 (20%)</td>
<td>21 (44%)</td>
</tr>
<tr>
<td>Annual income: $10,000 - $19,999</td>
<td>1 (10%)</td>
<td>10 (21%)</td>
</tr>
<tr>
<td>Annual income: $20,000 - $29,999</td>
<td>1 (10%)</td>
<td>6 (13%)</td>
</tr>
<tr>
<td>Annual income: $30,000 - $39,999</td>
<td>0</td>
<td>3 (6%)</td>
</tr>
<tr>
<td>Annual income: $40,000 - $49,999</td>
<td>0</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>Annual income: $50,000 - $59,999</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Annual income: $60,000 - $69,999</td>
<td>2 (20%)</td>
<td>0</td>
</tr>
<tr>
<td>Annual income: $70,000 - $79,999</td>
<td>1 (10%)</td>
<td></td>
</tr>
<tr>
<td>Annual income: $80,000 - $89,999</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Annual income: $90,000 - $99,999</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Annual income: $100,000 - $149,999</td>
<td>1 (10%)</td>
<td>0</td>
</tr>
<tr>
<td>Annual income: More than $150,000</td>
<td>1 (10%)</td>
<td>0</td>
</tr>
<tr>
<td>Annual income: Prefer not to disclose:</td>
<td>1 (10%)</td>
<td>3 (6%)</td>
</tr>
</tbody>
</table>

While we are neither lawyers nor legal experts, we note that by deidentifying participant responses before analysis and not maintaining a link to persistent identifiers, we believe we provide reasonable risk mitigation for participants.
Table 5.2: SDNS Providers Used by Participants. Service providers that participants indicated they had used, the total participants who indicated using them, and whether they respectively offered SDNS and/or VPN services. Demarcation with Yes* indicates that the provider previously offered this service, but no longer appeared to do so based on their website. Service providers demarcated with two asterisks (**) appear to no longer be in business.

<table>
<thead>
<tr>
<th>Service Provider</th>
<th>Total Used</th>
<th>SDNS</th>
<th>VPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NordVPN</td>
<td>42</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SmartDNSProxy</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PureVPN</td>
<td>9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SurfShark</td>
<td>9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>VPNSecure</td>
<td>6</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ExpressVPN</td>
<td>4</td>
<td>Yes*</td>
<td>Yes</td>
</tr>
<tr>
<td>Unblock-Us**</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HolaVPN</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>BulletVPN</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DNSFlex</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SmartyDNS</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Windscribe</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Blockless**</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CactusVPN</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>GetFlix</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IBVPN/IBDNS**</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>KutoVPN</td>
<td>1</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Mullvad</td>
<td>1</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>OperaVPN</td>
<td>1</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>ProtonVPN</td>
<td>1</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SimpleTelly</td>
<td>1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>StrongDNS</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>VPNUK</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

5.1.4 Study Limitations

SDNS is a relatively new offering, and although there appear to be a large number of SDNS users [41, 72], there are fewer discussion forums devoted to SDNS than there are VPNs, making it challenging to identify populations of SDNS users to recruit as participants. It is also possible that our recruitment efforts are affected by users’ wariness of sharing that they have used SDNS.

Due to the number of participants we were able to recruit, we cannot guarantee our participants make up a fully representative sample of SDNS users. As such, we can neither
assess the prevalence of the themes identified across all SDNS users, nor perform more in-depth quantitative analysis (e.g., correlation or regression studies).

Additionally, as we describe in more detail in §5.2, participants often struggled to distinguish between SDNS and VPN. This in turn made it harder to determine which of them had actually used SDNS. As we note in §5.1.2, some \( n = 7 \) participants claimed to have used SDNS, but upon further inspection were found not to have. Specifically, one participant failed attention checks, five participants listed providers that neither currently nor previously offered SDNS, and one respondent had free text responses indicating their exclusive use of their provider’s VPN offerings. However, free text responses from participants in this subgroup who passed all attention checks were still included in qualitative analysis, since their responses still indicated participants’ underlying confusion (see §5.2.4).

Despite these shortcomings, this study highlights many important themes amongst SDNS users’ perceptions, many of which, we believe, add valuable insight to the broader study of geoblock evasion, and how usage of evasion tools impacts the online security and privacy of their users.

5.2 Results

In this section, we describe the results of our study. We denote participants whose responses were only included in qualitative analysis with the letter Q. Participants whose responses were included in both quantitative and qualitative analysis are identified with the letter P. For the reader’s reference, we include our survey instruments and our codebooks in Appendices A and B respectively.

5.2.1 Motivation to Use SDNS

The vast majority \( n = 52 \) of participants noted they mainly used SDNS to access websites or other content to which their access would otherwise be restricted (M8). For most of these respondents \( n = 38 \), this meant bypassing server-side geoblocking, often \( n = 18 \) to use
a performance-sensitive service such as video streaming that required low latency and high bandwidth.

Interestingly, several \( n = 12 \) participants sought to use SDNS for protection it did not offer. These participants used SDNS for internet privacy \( n = 7 \), online anonymity \( n = 2 \) and online security \( n = 3 \). Amongst those seeking online privacy, two participants specifically mentioned limiting online tracking by corporations and government agencies. For example, P18 notes they specifically felt SDNS would help “...to minimise the collection of my data by companies who like spamming individuals with personalised ads.” In contrast, Q5, who sought to avoid government tracking, chose to use SDNS “...[to avoid] being tracked by the FBI,” implying potentially more serious consequences in the event that SDNS failed to sufficiently protect their online privacy.

Given the frequency with which SDNS is offered alongside VPN services, we asked participants why they specifically decided to use SDNS (M10). For many \( n = 21 \) participants, the answer lay in SDNS’ lower latency overhead, and its improved internet quality of service (QoS). Specifically, 18 participants noted experiencing lower latency in their internet browsing with SDNS and three participants noted it being a more lightweight and efficient solution to avoid geoblocking than VPNs. As P38 notes, increased network latency is especially noticeable when using online streaming services: “What motivated me to specifically use Smart DNS was the fact that it does not slow down my internet connection, more specifically, when using streaming services.” Two participants also note an additional reason for SDNS’ increased efficiency. As P43 explains, SDNS “…only forwards the necessary data for the geo unlock through the designated servers[.] …[This prevents] speed issues when using the service.”

For six participants the choice to specifically use SDNS was informed by its recommendation by someone they trusted. For most \( n = 5 \) of these participants, that person was a friend. However, for one participant (P40), the recommendation came in the form of an “…advertise[ment] by one of the content creators on youtube I watch regularly.” Among the
remaining participants, other frequently cited reasons for choosing to use SDNS included it being less expensive to use than VPN services \((n = 3)\) and participants’ belief that it more effectively evaded detection by content providers \((n = 3)\).

However, one group \((n = 5)\) of participants indicated that they did not know the difference between SDNS and other offerings. For example, P49 stated, “Marketing probably, because I dont know how difefent [sic] work VPN and Smart DNS[,]” This lack of understanding of how SDNS services operated was a common theme in many participants’ responses, as we describe in more detail in §5.2.4.
5.2.2 Selection of SDNS Provider

While most participants were able to distinguish between SDNS and VPN to some degree, we still observe a high overlap between the factors participants considered when choosing an SDNS provider, and those Khan et al. [75] note as VPN users' key motivations behind their choice of provider.

As shown in Figure 5.1, participants most strongly considered service price with 24 citing it as the top factor and 52 listing it as one of the top three factors considered overall. Their next most considered factors were the provider’s trustworthiness, with 12 citing it as the top factor and 35 citing it as one of the three factors, and provider’s ratings on review sites, 10 noting it as the top factor and 27 listing it among the three most important factors (M11).

5.2.3 Trustworthiness of Selected Providers

Given that trustworthiness of a provider was a common criterion among our participants when selecting their SDNS service, we also consider how participants formed their assessment of trustworthiness. When asked to explain their view on SDNS’s overall trustworthiness (M20), 16 participants spoke about positive factors that led them to believe their providers were trustworthy.⁸ Among these participants, seven cited providers’ projected image of security, six noted provider reviews, and four highlighted their provider’s overall reputation.

Among the seven participants who mention their providers’ projected image of security, most (n = 6) explain that their services provide anonymity and/or added security and privacy. Of those, two participants explicitly tout their provider’s use of safeguards such as data encryption, and “no logging” policies to protect their online security and privacy. As P13 (erroneously) states, “[a]ll of Smart DNS proxy servers are encrypted and secured. There are no logs, so all your traffic and data remains anonymous [sic] …” However, for P51, this image stems in part from SDNS providers’ general orientation towards security,

⁸We exclude the ten participants who cited observing that their service worked reliably, as well as the seven participants who stated they did not observe any indication of something bad happening when using the service, as it is unlikely these participants made either of these observations before they had actually chosen a provider.
noting “[t]he companies involved do seem to be primarily based around online security, so being trustworthy is fundamental to their businesses…”

For the four participants who considered their providers’ reputation, checking providers against additional criteria was generally necessary to convince them of their reputability. As other participants explained \((n = 3)\), not all SDNS providers can be trusted. Some providers, as P22 and P45 note, are scams that either “[claim] to offer a smart DNS [but] just charge you for nothing” (P45), or use the rouse of providing SDNS as a means to “…take your info” (P22). Among the criteria these four participants use to determine reputation are user reviews and recommendations \((n = 1)\), providers’ years of experience offering their services \((n = 1)\), and their overall track record \((n = 1)\).

As shown in Figure 5.2, once participants had chosen an SDNS provider, most of them \((n = 48)\) found it trustworthy overall \((M_{18})\).

**Breaking Down Trust** We next consider to what degree participants who use SDNS services trust their providers to do certain actions \((M_{21})\). Unsurprisingly, far more partici-
pants \((n = 54)\) indicated that they trusted their provider to bypass geoblocking—one of the primary functions of SDNS services—while only four indicated that they did not.

However, when asked about their willingness to trust their chosen provider to respect and protect their privacy, more participants indicated hesitancy to do so. Specifically, when asked if they trusted their SDNS providers to abide by their privacy policy, fewer participants \((n = 40)\) indicated that they did (relative to the 54 who trusted their ability to bypass geoblocks).
While the majority of participants still trusted their SDNS provider to respect and safeguard their privacy and to use industry best practices in security, many (but not most) participants were hesitant to do so. As Figure 5.3 further illustrates, only 36 participants trust their provider to respect and safeguard their privacy while 35 trust them to use industry best practices in security (M21).

When asked to explain why they trusted these services (M20), participants mainly cited that the service reliably worked well (n = 10) and that nothing bad had happened, or that they lacked a reason to distrust their providers (n = 7). That is, many participants acknowledged that by choosing to use SDNS, they were likely taking a risk with their online security and/or privacy. Despite this, these respondents indicated not noticing any evidence of their SDNS service providers harming them. As P33 explains, “. . . Personally, I do trust these services. They must have the best kinds of security measures to protect their user’s data, so that their whole client base doesn’t get backstabbed by [the] company [to whom] they are giving basically all their network traffic information.”

The remaining respondents (n = 27) implied they still had (varying degrees of) lingering reservations about their providers. Chief among their concerns were how SDNS providers used their personal data (n = 9). As P31 notes, many SDNS providers lack transparency about how they handle their users’ data: “I think i [sic] can trust they service, but I’m not so sure how my personal information is used.” Similarly, Q1 notes, “Some may very well make significant money exchanging your data.”

5.2.4 Understanding of SDNS Functionality

As indicated by their motivations to use SDNS (see §5.2.1), most participants understood that SDNS providers enabled the bypassing of geoblocks. However, many participants struggled to distinguish between SDNS and VPN. In particular, participants largely mis-attributed VPN functionality to SDNS. When asked why they specifically chose to use SDNS rather than a VPN or any other tool capable of bypassing geoblocks (M10), four
respondents explicitly remarked that they did not know the differences between SDNS and VPNs, and one participant stated that they were unfamiliar with other circumvention options.

Confusion over Prior SDNS Use Many participants were not able to accurately distinguish their usage of VPNs from that of SDNS. Specifically, five participants who stated they had experience using SDNS either as a current SDNS user or as a previous one had in fact confused their usage of a VPN for that of SDNS. Among these respondents, four stated they had used SDNS provided by providers that only offered VPN services, and one provided open responses that strongly indicated they had exclusively used the VPN.

Among participants who did distinguish between SDNS and VPNs (and other circumvention options), many still showed confusion, or misconceptions about how SDNS works.
Visibility of Browsing to SDNS Provider

Sees my browsing once set up

Sees my browsing even if no blocked sites visited

Figure 5.5: Perceived Visibility of Internet Browsing to SDNS Provider. Participant beliefs about the visibility of their browsing to their SDNS provider (M6).

As shown in Figure 5.4, much of this disconnect concentrated around the services’ routing and proxying behaviors. Specifically, participants’ conceptualizations of SDNS systems did not seem to encompass how and when SDNS routes their traffic to proxies.

5.2.5 Understanding of SDNS’ Impacts on Security and Privacy

We found that for many participants, there was a major disconnect between their mental models of how SDNS operates and actual SDNS functionality. For example, when asked about how SDNS proxies their internet traffic (M5), less than half ($n = 28$) of participants stated (correctly) that SDNS proxied only some of their internet traffic rather than all of it.

More troubling, participants also tended to underestimate SDNS services’ ability to track their users’ browser behavior. Because all DNS requests are routed through SDNS services’ DNS resolvers, SDNS services have significant access to browsing behavior. However, when asked whether their SDNS provider could determine which websites they visited (M6), less
than half of the participants ($n = 26$) correctly indicated that their SDNS provider could see their internet browsing behavior, as shown in Figure 5.5. Even fewer participants ($n = 20$) understood that SDNS services can observe internet behavior even if no blocked sites were visited.

More broadly, participants by and large appeared to put undue confidence in the security and privacy offered by SDNS services. As shown in Figure 5.6, the vast majority of participants ($n = 48$) at least somewhat agreed that SDNS improved their online security when browsing the internet ($M_1$). Even more participants ($n = 50$) indicated that SDNS helped protect their privacy when browsing the internet ($M_2$).

When asked about SDNS’s security and privacy properties in more detail (see Figure 5.7), slightly less than half of participants ($n = 26$) indicated that SDNS encrypts web traffic (in actuality, it does not). Slightly fewer ($n = 25$) indicated that SDNS hides their browsing behavior from their internet service provider (ISP); this too is false, since (1) SDNS services do not encrypt traffic and (2) ISPs can observe unencrypted DNS requests to SDNS resolvers.
Additionally, through participants’ free text responses (M4, M8, M10, M20), we observe several pervasive misconceptions about the security provided by SDNS services. Chief among these were SDNS gives its users (some degree of) anonymity online \( (n = 11) \), and using SDNS helps prevent tracking of one’s online behavior \( (n = 9) \). Amongst the parties from whom respondents claimed SDNS protected against tracking, were advertisers or corporations \( (n = 3) \), governments \( (n = 2) \), any third party \( (n = 2) \), malicious entities \( (n = 1) \), and participants’ ISPs \( (n = 1) \). For P36, the perception that using SDNS would offer anonymity and privacy also served as a large motivation for their decision to use SDNS and not a different service (M10): ‘The ability to feel free knowing that barely noone [sic] can identify
me. Obviously this isn’t as secure as using .onion, but it narrows down the possibilities for people and organisations to know who I am.”

Although not nearly as pervasive, we observe two other noteworthy misconceptions about the security and privacy provided by SDNS. These include perceptions that SDNS was more safe or secure than a VPN ($n = 3$) and that using SDNS helps mitigate the risks of hacking or malware infection ($n = 2$).

### 5.2.6 Ethics and Legality of Using SDNS

**Ethics of using SDNS** As shown in Figure 5.8, most participants ($n = 42$) agreed that using SDNS is ethical. In their reasoning, they cite several factors including the belief that geoblocking is unethical ($n = 27$), and that they were still paying the content providers whose content they accessed ($n = 5$).
Amongst the 27 participants who justify their use of SDNS, eight cited their belief in a free and open internet. P32 explained that “the internet has evolved as a virtual world without borders, and it is right that it remains so.” Some of these participants \((n = 3)\) described geoblocking as a form of discrimination. For example, P11 and P18 stated:

I believe in free use and free access i [sic] believe it to be unfair and bordering racism if you don’t provide users with the same content that you would other if there is no legitimate reason otherwise. (P11)

I pay each and every month to get access so why am I limited to content just because of my geographic region? I think it’s very unfair because I feel like I’m being punished for living in a certain country. (P18)

Others \((n = 5)\) remarked that they were not pirating the content and that they were still paying to access it. For example, P37 noted:

I used Smart DNS to access streaming services that are unavailable in my country. I still paid for them so I think it’s a better option than piracy because the company got their money.

Finally, four participants stressed that they are not causing harm by evading geoblocks. P57 exemplifies this very well: “Why would something like this be unethical. I’m not doing any harm to anyone. I’m just watching Netflix.”

Among the participants whose answers indicated they did not believe bypassing geoblocks using SDNS was ethical, one participant (P49) noted regional copyright and usage rights issues:

This is hard to explain, its [sic] just comes to tv rights. Maybe netflix has bought the rights for a series, lets say game of thrones but only for america, because in other countries another company has the rights. Well if you are entering and watching game of thrones in that other country u should do it through the company that has the rights who is paying for them.
Given SDNS is used frequently (if not primarily) to access online streaming services, this is noteworthy, as compliance with licensing restrictions is likely the main reason why many streaming online content providers perform geoblocking.

**Legality of using SDNS** While a majority of participants ($n = 31$) also believe that using SDNS is legal, this majority is slimmer than that for the ethics of using the service. This seems to reflect participants' widespread confusion and lack of awareness of how their local laws address geoblock bypassing, if at all. When asked to explain their reasoning, participants' answers varied. Several participants ($n = 17$) stated that they do not know whether or not it is legal, and another four respondents indicated that its legal status depends on the user’s local laws ($n = 3$) or on the purpose of use ($n = 1$). In more detail, 12 participants point out that there is a wide regional variance in local laws governing the usage of tools such as SDNS to bypass geoblocking. In fact, most ($n = 7$) of these respondents state that there likely is no local law prohibiting them from using SDNS.

Other participants ($n = 2$) point out that if geoblock bypass via SDNS is illegal, the laws banning it are not enforced. As Q2 notes, “I think it is legal, but if its not noone is going to come to your house, ’cause a lot of people do worst things in the internet . . .”

As P1 described, this lack of clarity is only magnified by the prevalence of mainstream ads that serve to normalize SDNS usage: “I’m not sure but I would guess it has to be legal for them to be so popular, and advertised on mainstream media (like YouTube).” In some cases, this normalization is also reflected on the information published on various SDNS provider websites. For example, in at least three SDNS providers’ FAQ pages, the services argue that their services are legal since it is lawful to both change one’s DNS settings and use a proxy server [62, 76, 92]; we lack the expertise to authoritatively assess the legal persuasiveness of these arguments.
5.3 Discussion

A troubling finding of our study is that most of our participants did not consider the consequences to their security and privacy of using SDNS services. As illustrated by participants’ responses, many were not cognizant of the types of vulnerabilities to which they are susceptible when using SDNS. Worse still, many participants were prone to putting undue trust in the protections their SDNS providers purported to offer. This overall dynamic only serves to compound the privacy threat SDNS poses to its average user.

Given the nature of the privacy vulnerabilities identified in Chapter 4, SDNS users are unlikely to recognize whether their privacy has been breached. Specifically, the attacks we identify do not give their victims any indications that anything went awry. As such, participants’ widespread belief that nothing nefarious has happened due to their SDNS usage may be inaccurate.

Since, as noted in Chapter 4, many of the privacy vulnerabilities associated with SDNS usage are inherent to these systems’ architecture, there is no straightforward means to adequately remedy them or mitigate the risks they pose while using SDNS. As such, the best way a user could protect themselves against these threats to their privacy would be to stop using SDNS altogether. To continue evading geoblocking, users could switch to a geoblock evasion tool with more robust security, such as a VPN, or Tor. For participants who indicated using SDNS because they (1) were unfamiliar with other geoblock evasion tools, (2) did not know the difference between SDNS and other options for bypassing geoblocks, or (3) had misconceptions about SDNS’ impact on their online security and privacy, user education may be beneficial. This education would need to explain the security and privacy risks of using SDNS, and to address common user misconceptions about these systems. In particular, educators would need to address common user misconceptions about SDNS’ impact on their online security and privacy, and to stress that (1) SDNS usage inherently opens users up to privacy risks to which they would not otherwise be vulnerable; and (2) that a user whose privacy has been breached often receives no indication that this occurred.
One means by which this could be achieved is through the publication of accurate and easily accessible information about SDNS systems, how using them impacts users’ online security and privacy. In the case of selecting a trustworthy VPN provider, Khan et al. recommend providing easy access to unbiased, thorough and peer-reviewed information about the most prominent service providers [75]. They argue that doing so would help inform VPN users’ decisions and encourage them to make more secure choices of provider. Such advice may also be applicable to SDNS users. As such, we recommend the publication of online guides or articles describing the risks associated with using SDNS. Like many existing VPN guides including, but not limited to those published by the EFF [45, 46], Consumer Reports [59], and others, these guides should be widely available and written to be accessible to individuals lacking technical backgrounds.

To redirect willing SDNS users to a more secure means of geoblock evasion, education would also need to provide guidance on how to select a trustworthy and secure alternative to SDNS. Such guidance would need to either include basic descriptions of these options including their respective capabilities and limitations, or link to existing sources with this information.

However, while education is likely to be beneficial to many SDNS users, it would not be reasonable to expect it to consistently empower/influence users to stop using SDNS. As noted in §5.2.4, many participants chose to use SDNS in lieu of VPN due to its increased usability. Specifically, many participants mention the seamlessness of SDNS’ service, its increased success at evading detection by content providers, and it not requiring custom software. Previous research by Kang et al. finds that internet users are often deterred from taking actions to protect their privacy when they perceive they would have to sacrifice convenience do so, or that the software tools aimed at protecting their privacy have poor usability [73]. Ruoti et al. expand on this further, noting that, when determining which security and privacy measures to adopt, users opt for measures that will not impede their ability to use the internet [109]. Given our findings indicating many SDNS users view geoblocking as an
unethical practice, a form of censorship and discrimination, and as being directly in conflict with their belief in the open internet, these users may see successful geoblock evasion as a prerequisite for or core component of their ability to use the Internet. Therefore, persuading these SDNS users to adopt VPNs instead would likely require a VPN service that boasted similar, if not better, usability than the SDNS services that are currently available. In the case where these SDNS users find the VPN falls short of offering comparable usability to SDNS, they may decide to continue using SDNS services despite the risks they pose.

5.4 Summary

Beyond the scope of SDNS, our findings in this chapter provide a bleak outlook for the online security and privacy of users of geoblock evasion tools in the absence of additional user education.

More generally, when it comes to geoblocking, some users’ lack of awareness may be systemic. As illustrated by many participants’ negative perceptions of geoblocking and their beliefs that geoblocking was unethical it is not clear whether most Internet users understand why geoblocking is commonly used, or even whether there is a widespread consensus over when geoblocking is acceptable, if at all.

It is worth noting, that cases where a user’s beliefs and/or perceptions of geoblocking’s ethics or legality do not match those outlined within their local laws could put them at risk. Put differently, a user’s belief that using an unblocker is ethically or legally justified to protect their access to the free and open Internet has the potential of significant legal consequences in cases where the laws to which they are subject view the practice differently.

Beyond awareness of unblockers’ legality, those who choose to use these tools often lack the means to make a well-informed choice of service or provider. Similar to Khan et al.’s observations, we note, there is currently very limited reliable guidance on the potential security risks and benefits of using each of the different geoblock evasion tools available.

---

9 We once again reiterate that we are neither lawyers, nor legal scholars and are unqualified to determine the legal status of using any means of geoblock circumvention.
Worse still, this dearth of reliable information extends to guidance on how to ascertain the trustworthiness of a given service purveyor.

Given their lack of means for assessing these, Internet users who decide to turn to an unblocker often end up relying on service reviews, which as Khan et al. find, and we also observe, are heavily biased due to many unblockers’ wide networks of affiliate programs, in which authors of review sites are paid a commission for each new customer they attract to their affiliated unblockers’ service [75]. As a result, many users are therefore woefully ill-equipped to choose a trustworthy unblocker, let alone recognize how or when their use of an unblocker could put their security or privacy at risk.

Therefore, in the face of pervasive demand for ways to evade geoblocking, similarly widespread user education on both geoblocking and on unblockers is imperative to safeguarding Internet users’ online security and privacy.
This thesis describes a systemic study of geoblocking and its impacts on both Internet ecosystems and users whose online access it has restricted.

Similar to previous studies’ observations [17], we reiterate that major websites’ increased reliance on CDNs has effectively centralized where Internet connection requests are routed and the infrastructure used to respond to them.

We note that the impacts of this have likely increased in magnitude recently due to the continued expansion of websites’ general reliance on CDNs and that they will continue to do so in light of the ongoing increase in available IP-geolocation ground truth data. Specifically, this added data has become available through the introduction of geofeeds in RFC 8805 [77] and the expansion of RPSL to allow the WHOIS registries to become a central directory through which these feeds can be located and authenticated as defined in RFC 9092 [103]. Therefore, we measure the growth of geofeed registration and publication within the WHOIS registries over the course of 17 months and assess their impact on two of the commercial IP-geolocation services most widely used by popular CDNs on the Internet.

As we discuss in more detail in Chapter 3, we find strong evidence indicating commercial IP-geolocation providers likely consult geofeed ground truth data and incorporate it into their underlying DBs when it is available. However, while we find substantial evidence of commercial IP-geolocation services’ reliance on available geofeed data, we find that its publication has not become sufficiently widespread to meaningfully affect their overall rates of accuracy. We therefore conclude that should a substantial proportion of network operators opt-into registering and publishing geofeeds for their networks, the resulting data will
have a profound impact on commercial IP-geolocation services’ accuracy, which in turn will significantly affect geoblocking and other website customizations based on users’ respective vantage points.

In particular, our study of their proliferation indicates that to date most of its adoption is concentrated in countries that are more industrialized and have a larger Internet presence and that countries with smaller Internet presence are often underrepresented within the publicly available geofeed results. As noted in § 3.6, this serves as an important warning. Specifically, should this trend persist as geofeed publication and commercial IP-geolocation services’ reliance therein become more commonplace, it could result in commercial IP-geolocation services being significantly less accurate when locating vantage points situated in countries with less Internet infrastructure, which in turn could lead to further deterioration of website accessibility and Internet QoS in these locations. Since prior work in the literature has found significant discrepancies in website accessibility and Internet QoS between vantage points with robust Internet presence and those situated in countries with less Internet infrastructure [116], the potential widening of these gaps pose significant threats to a free and open Internet.

To better understand some of the broader implications of this potential increase in geofiltering and the proliferation of its evasion, we measure the ecosystem, architecture and security and privacy properties of a geoblock evasion system with limited prior study. Overall, we find that beyond limiting website accessibility, geoblocking has secondary but equally profound repercussions for end users’ online security and privacy. In the case of SDNS, we observe that these systems’ usage for geoblock evasion pose significant risks to its users’ online privacy. Among the threats to user privacy which we find SDNS systems are susceptible are enumeration of SDNS users’ IP addresses by third parties or supported channel operators through the spoofed DNS resolution requests, users’ increased vulnerability to eavesdropping due to their traffic traversing longer network paths, and their increased exposure to analytics of their Internet browsing behavior. We observe that the susceptibilities to
attack we discover either stem from faulty system configurations, or are due to architectural weaknesses that are inherent to SDNS systems’ setup.

However, upon further study of SDNS users’ perceptions of these services and their providers, we encounter a stark contrast with our system measurement observations. Specifically, we find that many users think highly of their SDNS providers, and in some cases (incorrectly) perceive SDNS as a privacy enhancing technology. Among the many contributing factors to users’ perceptions, we note that many users hold geoblocking in negative esteem, and in some cases wish to circumvent it as a result. When coupled with the widespread publication of potentially misleading, false or incomplete information about geoblock evasion tools’ security and privacy properties on provisioners’ websites and on those of their affiliates, this renders disaffected Internet users susceptible to putting undue trust in their chosen geoblock evasion tools or in their provisioners.

While far from conclusive, these findings hint at what could be an increasingly bleak picture for Internet users in countries with less Internet infrastructure. In this scenario, these users’ Internet accessibility and QoS would continue to deteriorate alongside network operators’ expanding opt-in to publishing geofeeds, and commercial IP-geolocation providers’ increasing reliance on their results. In the worst case scenario, this deterioration could precipitate users’ increased adoption of geoblock evasion tools like SDNS in regions where Internet accessibility and QoS has suffered. Given our findings of widespread misconceptions about the security and privacy properties of these tools amongst their existing users, we posit that users who turned to these tools would have similar misconceptions about them, and as a result would be all the more susceptible to the threats these tools pose to their online security and privacy.

Given the gravity of this outcome, more work is needed to better understand the likelihood of these events and whether they are causally related. To that end, we recommend future work focus on the following areas:
1. The continued measurement of geofeed expansion and of commercial IP-geolocation services’ use of/reliance on them as ground truth data.

2. Continued Internet accessibility and QoS measurements – particularly from vantage points in countries whose Internet presence is more modest and to which the fewest geofeed IP addresses are geolocated.

3. Geoblocking mechanisms and their prevalence, particularly those used for partial geoblocking, which to the best of our knowledge has not been studied in the open literature.

These findings will provide clearer insights into future Internet accessibility around the world, and ideally identify the ways in which it can be protected as the Internet continues to evolve.
APPENDIX A

SURVEY INSTRUMENTS

The following question ordering reflects how this survey was presented to Reddit participants. Participants who took the survey through Prolific were given questions S1 through S9, and D1 through D6 as a screening survey, and completed M1 through M31 as the main survey.

S1 How familiar are you with Smart DNS services?
- Not at all familiar
- Moderately familiar
- Slightly familiar
- Somewhat familiar
- Extremely familiar

S2 What are Smart DNS services primarily used for?
Answer: ________________________________

S3 Do you currently use Smart DNS, or have you done so in the past?
- Yes
- No
- I'm not sure

S4 Are you seriously considering using Smart DNS within the near future?
- Yes
- No
- I prefer not to answer

* S4 only shown if answer to S3 is not “Yes”*

S5 Which SmartDNS services have you used before or are you seriously considering using? (Select all that apply)
- AceVPN
- Blockless
- BulletVPN
- CactusVPN
- DNSFlex
- GetFlix
- HideIPVPN
- Invisible Browsing (IBVPN/IBDNS)
- IronSocket
- KeepSolid
- Le-VPN
- NordVPN
- Overplay
- PureVPN
- SimpleTelly
- SmartDNSProxy
- SmartyDNS
- StrongDNS
- SurfShark
- TrickByte
- TVWhenAway
- Uflix
- Unblock-Us
- Unilocator
- VPNSecure
- VPUK
- Other: ________________________________

* denotes an exclusive answer.

* S5 only shown if “Yes” answered to S3 or “Yes” answered to S4.

S6 Which of the following types of Smart DNS accounts have you had or used (including any accounts you currently have/use)? (Select all that apply.)
- Free Trial
- Paid Subscription
- I Used Someone else’s account
- I am unsure*
- I have never had nor used a Smart DNS account*

* denotes an exclusive answer.

Now we are going to ask you questions about the **Domain Name System (DNS)** which is different from **Smart DNS**.
S7  How familiar are you with the Domain Name System (DNS)?
○ Extremely Familiar
○ Very Familiar
○ Moderately Familiar
○ Slightly Familiar
○ Not Familiar at all

S8  Do you know how the Domain Name System (DNS) works?
○ I definitely know
○ I somewhat know
○ I’m not sure I know
○ I definitely do not know

S9  To the best of your knowledge, explain how DNS works by describing the steps taken by your computer when you navigate to a website like http://www.example.com
Answer:

M1  Smart DNS provides additional security when browsing the internet.
○ Strongly agree
○ Agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Disagree
○ Strongly disagree

M2  Smart DNS provides additional privacy when browsing the internet.
○ Strongly agree
○ Agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Disagree
○ Strongly disagree

M3  Using Smart DNS is a risk to my security and privacy.
○ Strongly agree
○ Agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Disagree
○ Strongly disagree

M4  Please explain why you think Smart DNS can affect your security and privacy in the way(s) you indicated.
Answer:

M5  Note whether you agree or disagree with each of the following statements about Smart DNS.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Smart DNS encrypts my web traffic.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) Smart DNS slows down my Internet connection.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iii) Smart DNS Can make my Internet traffic go through a proxy server.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iv) Smart DNS lets me access websites and/or content that I otherwise couldn’t access.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(v) Smart DNS speeds up my Internet connection.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(vi) It is difficult for a website to determine if I am using Smart DNS to access it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(vii) When I use Smart DNS, some sites are accessed via Smart DNS’ proxies, while others are not.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Note whether you agree or disagree with each of the following statements about Smart DNS.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(ii) Once I have set up Smart DNS on my computer, my Smart DNS provider can see which websites my computer visits.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iii) Once I have set up Smart DNS on my computer, my Smart DNS provider can see which websites my computer visits even if I never use it to visit blocked websites.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iv) Smart DNS makes it more difficult for a blocked website to determine my IP address.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(v) If I access a blocked website using Smart DNS, the website will not be able to tell I am using Smart DNS.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M7 It is worthwhile to put in the effort to use Smart DNS.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

M8 What was your main goal in using a Smart DNS service?

Answer: ________________________________

M9 Which countries’ websites do/did you most often access using Smart DNS services? To select multiple options hold down the Control (Command on Mac) and click the answers you would like to select. To deselect a single answer hold down the Control key (Command on Mac) and click the option you would like to deselect.

*Dropdown list of countries*

M10 Like other services, Smart DNS services come with a specific set of strengths and weaknesses. Given that Smart DNS services are not the only offerings capable of unblocking geo-fenced websites (for example, some VPN services can do this as well), what motivated you to specifically use Smart DNS?

Answer: ________________________________

M11 When you chose a Smart DNS service, which of the following factors did you consider most relevant? (If you plan to use Smart DNS in the near future, which of the factors are you most strongly considering?) Select up to three factors, and rank them by entering 1-3 in the text box to their left, where 1 is the most relevant item selected.

- Price
- Service’s rating on a review site
- Offered channels
- Additional service offerings
- Provider’s Trustworthiness
- Customer support
- Company location (e.g., based in US, UK, etc.)
- They offered a free trial
- I just used the first one I found
- Other (please specify): ________________________________

M12 Does the Smart DNS provider you use, used, or plan to use, offer any services in addition to Smart DNS?

- Yes
- No
- I’m Unsure
M13 and M14 are only displayed if participant answered "Yes" to M12.

M13 When you signed up for your provider’s service(s) were you primarily looking to use their Smart DNS?
- Yes
- No
- I’m Unsure

M14 Which other services does/did your Smart DNS provider offer? (Select all that apply)
- I’m not sure*
- VPN
- P2P Torrent Support
- Ad Blocking
- Firewalls
- Other (please specify):________________________________________

* denotes exclusive answer

M15 is only displayed if participant answers "Yes" to M12 and does not select "I’m not sure" on M14.

M15 Which, if any, of these services have you used, or do you plan on using? (Select all that apply)
[Selected choices from M14]
- I did not use, nor plan on using any other services*

* denotes exclusive answer.

M16 When was the last time you set Smart DNS on a device?
- in the past few months
- in the past few days
- I have not set up Smart DNS on a device

M17 Have you set up Smart DNS on the device you are using to complete this survey?
- Definitely yes
- Probably yes
- Might or might not
- Probably not
- Definitely not

M18 Is Smart DNS currently enabled on the device you are using to complete this survey?
- Yes
- No
- I’m Unsure

M19 How trustworthy do you find these services overall?
- Very Trustworthy
- Slightly Trustworthy
- Neither Trustworthy nor Untrustworthy
- Slightly Untrustworthy
- Very Untrustworthy

M20 Please describe your view on the overall trustworthiness of Smart DNS.
Answer: _________________________________________________________

M21 To what extent do you trust your Smart DNS provider to:

<table>
<thead>
<tr>
<th></th>
<th>Definitely Trust</th>
<th>Probably Trust</th>
<th>May or May Not Trust</th>
<th>Probably Do Not Trust</th>
<th>Definitely Do Not Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Allow you to bypass blocking as advertised?</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(ii) Abide by its privacy policy?</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(iii) Respect your personal data and keep it private?</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(iv) Use industry best practices in security?</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

M22 Based on your experience, how often does Smart DNS successfully allow access to blocked content?
- Always
- Sometimes
- Most of the Time
- Never
- About Half of the time

M23 Have you ever been blocked by a content provider because you were using Smart DNS?
M24 How do you think they determined you were using Smart DNS?
Answer: ________________________________

M25 How frequently are you blocked for this reason?
○ Always
○ Most of the Time
○ About Half of the time
○ Sometimes
○ Never

M26 How easy do you think it would be for content providers to determine if you were using Smart DNS?
○ Extremely easy
○ Moderately easy
○ Slightly easy
○ Neither easy nor difficult
○ Slightly difficult
○ Moderately difficult
○ Extremely difficult

Ethics and Legality Because Smart DNS allows you to access content that would normally be unavailable in your geographic region, there may be disagreement among users regarding the ethics and legality of using Smart DNS. Based on your opinions, state how much you agree, or disagree with each of the following statements. After each statement, you will be asked to explain your response.

M27 Using Smart DNS to access content outside my geographic region is ethical.
○ Strongly agree
○ Agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Disagree
○ Strongly disagree

M28 Please explain your prior response:
Answer: ________________________________

M29 Using Smart DNS to access content outside my geographic region is legal.
○ Strongly agree
○ Agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Disagree
○ Strongly disagree

M30 Please explain your prior response:
Answer: ________________________________

D1 What is your age?
○ 18 – 24
○ 25 – 34
○ 35 – 44
○ 45 – 54
○ 55 – 64
○ 65 – 74
○ 75 – 84
○ 85 or older
○ I prefer not to disclose

D2 What is the gender to which you most closely identify?
○ Male
○ Female
○ Non-binary
○ I prefer to self-describe
○ I prefer not to answer

D3 What is the highest level of school you have completed or the highest degree you have received?
○ Less than a High School Degree
○ High school graduate (high school diploma or equivalent including GED)
○ Some college but no degree
○ 2 year degree (e.g. Associate degree in college or trade degree)
○ Bachelor’s degree in college (4-year)
○ Master’s degree
○ Professional degree (e.g., J.D., M.D.)
○ Doctoral degree (PhD)
○ I Prefer not to answer

D4  In which country do you currently reside?
    *Dropdown list of countries*

D5  What is your nationality?
    *Dropdown list of countries*

D6  What is your annual income in US Dollars?
    ○ Less than $10,000
    ○ $10,000 – $19,999
    ○ $20,000 – $29,999
    ○ $30,000 – $39,999
    ○ $40,000 – $49,999
    ○ $50,000 – $59,999
    ○ $60,000 – $69,999
    ○ $70,000 – $79,999
    ○ $80,000 – $89,999
    ○ $90,000 – $99,999
    ○ $100,000 – $149,999
    ○ More than $150,000
    ○ I prefer not to disclose
**Appendix B**

**Codebooks**

**S2: What are Smart DNS services primarily used for?**  
(Cohen’s $\kappa = 0.77$)

- unblock (36)
  - geo-restricted (27)
  - websites (4)
  - restricted (2)
  - age-restricted (1)

- access (20)
  - geo-restricted (19)

- change/mask-ip (6)
  - hide/spoof-geographic-location (2)

- hide/spoof-geographic-location (6)

- similar-to-vpn (6)

- using-dns (2)

- anonymity-on-Internet (1)

- proxy (1)

- N/A (1)

**S9 How DNS works**  
(Cohen’s $\kappa = 0.87$)

- maps-website-to-ip (17)
  - definition (11)
  - describes-steps (4)
  - finds-ip
    * describes-steps (1)

- maps-domain-to-ip (15)
  - definition (5)
  - describes-steps (1)

- query-recursive-dns (12)
  - describes-steps-detailed (12)

- sdns (8)
  - proxy (4)
M4: Please explain why you think Smart DNS can affect your security and privacy in the way(s) you indicated. (Cohen’s $\kappa = 0.76$)

- positive-impact (34)
  - limit/prevent-tracking (9)
    * by-advertisers (1)
    * by-government (1)
  - hides/masks-ip (8)
  - hides/masks-location (6)
  - anonymity (6)
  - security (5)
  - unblock (3)
  - hides-information (1)
  - verify-domain-validity (1)

- limited/no-impact (16)
  - doesn’t-hide/mask-ip (5)
  - no-encryption (4)
  - if-honest/known-provider (2)
  - metadata-still-accessible (2)
  - not-everything-proxyed (1)
  - not-intended-for-security (1)
  - police/government-subpeona-data (1)

- negative-impact (9)
  - cannot-trust-sdns/proxy (2)
  - increased-user-attack-surface (2)
  - unblocks-dangerous-websites (2)
  - risk-of-provider-security-breach (1)
  - risk-of-revealing-ip (1)
  - sdns-learns-more (1)

- depends (1)

- unsure (1)

- failed-to-answer (1)
  - don’t-know (1)
M8: What was your main goal in using a Smart DNS service? (Cohen’s $\kappa = 0.79$)

- access (43)
  - geo-restricted (30)
    - speed-sensitive (1)
  - restricted (8)
  - Netflix (3)
  - government-restricted (1)
- video-stream (12)
  - geo-restricted (9)
    - Netflix (5)
  - restricted (2)
- privacy (5)
  - mainly-use-vpn (1)
- security (3)
- unblock (3)
  - geo-restricted (1)
  - restricted (1)
- anonymity (2)
- avoid/limit-online-tracking (2)
  - by-government (1)
- free-trial (1)
- proxy (1)
  - decrease-latency (1)
- spoof-location (1)
- torrent-download (1)
- videogame-connectivity (1)

M10: Like other services, Smart DNS services come with a specific set of strengths and weaknesses. Given that Smart DNS services are not the only offerings capable of unblocking geo-fenced websites (for example, some VPN services can do this as well), what motivated you to specifically use Smart DNS? (Cohen’s $\kappa = 0.76$)

- lower-latency (18)
- unclear (7)
- recommendation/endorsement (6)
  - by-friend (5)
  - by-youtuber (1)
- don’t-know-difference (4)
• better (3)
• cheaper-than-vpn (3)
• more-effective-unblocking (3)
• mainly-use-vpn (2)
• proxies-minimum-to-unblock (2)
• safer-than-vpn (2)
• security (2)
• use-both (2)
• ad-blocking (1)
• added-safety (1)
• anonymity (1)
• curiosity (1)
• easier-to-use (1)
• efficiency (1)
• less-used (1)
• more-accessible (1)
• more-secure-than-vpn (1)
• more-stable-connection (1)
• no-software-needed (1)
• popularity (1)
• privacy-features (1)
• trustworthy (1)
• unfamiliar-with-alternatives (1)
• used-someone-else’s (1)

M21: Please describe your view on the overall trustworthiness of Smart DNS. (Cohen’s χ = 0.81)

• positive (33)
  – works-well (10)
  – nothing-bad (7)
  – good-reviews (3)
  – adds-security (3)
  – was-recommended (2)
  – anonymous (1)
  – caveat (1)
    * lack-transparency (1)
  – security-focused-providers (1)
  – trust-security-mechanisms (1)
• negative (12)
  – lack-transparency (5)
    * about-data-privacy (1)
  – assume-untrustworthy (2)
  – may-sell-data (1)
  – access-personal-info (1)
  – used-to-break-rules (1)
  – for-profit (1)
• positive-contingent (11)
  – provider-reputation (4)
  – good-reviews (3)
  – info-you-share (2)
  – most-are-legitimate (1)
  – paid-or-free (1)
  – provider (1)
  – provider-size (1)
• ambivalent-contingent (4)
  – provider (3)
  – paid-or-free (1)
    – provider-professionalism (1)
• ambivalent (2)
  – may-sell-data (1)
  – no-opinion (1)
• always-some-risk (1)
  – change-internet-path (1)
• unclear (1)

M25: How do you think they determined you were using Smart DNS? (Cohen’s $\kappa = 1.0$)

• don’t-know (3)
• ip-checks (3)
• location-changes (3)
• sdns-leaked-info (2)
• browser-cache (1)
• faulty-usage (1)
• location-data (1)
M29: Explain answer - Why using SDNS to access content blocked in my geographic region is/is not ethical  
(Cohen’s $\kappa = 0.85$)

- ethical (43)
  - against-geofences (27)
  - free-open-Internet (8)
  - against-censorship (5)
  - paying-for-content (5)
  - not-harming (4)
  - geofence-is-discrimination (3)
  - available-to-others (1)
  - gives-info-access (1)
  - not-available-otherwise (1)
  - not-disabling-blocks (1)
  - not-for-illegal-behavior (1)
  - travel-restrictions (1)

- unethical (7)
  - blocked-for-a-reason (2)
  - bypassing-blocks (2)
  - lying (1)
  - not-paying (1)
  - violates-geographic-TV-rights (1)

- don’t-know (5)

- contingent (2)
  - region (1)
  - use-of-access (1)

- ethical-contingent (2)
  - not-for-illegal-behavior (1)
  - reason-for-content-block (1)

- not-my-problem (1)

M31: Explain answer - Why using SDNS to access content blocked in my geographic region is/is not legal  
(Cohen’s $\kappa = 0.77$)

- legal (31)
  - no-local-law-against (7)
  - not-using-for-criminal-activities (3)
  - nothing-wrong (2)
  - paying-for-unblocked-service (2)
  - blockers-are-the-problem (2)
  - no-harm-done (2)
  - no-law-against (2)
  - available-in-origin-country (1)
  - mainstream-ads (1)
  - not-blocked-by-government (1)
- others-do-worse (1)
- people-do-it (1)
- privacy-my-right (1)
- ToS-aren’t-laws (1)

- don’t-know (17)

- not-legal (13)
  - blocked-for-a-reason (5)
  - authorities-are-blockers (1)

- contingent (4)
  - user’s-local-laws (3)
  - use-purpose (1)
Bibliography


