View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Harvard University - DASH

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

2010 Circumvention Tool Usage Report

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation	Hal Roberts, Ethan Zuckerman, Jillian York, Robert Faris, John Palfrey, 2010 Circumvention Tool Usage Report (2010).
Accessed	February 19, 2015 1:45:56 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:5366727
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

2010 Circumvention Tool Usage Report

Hal Roberts, Ethan Zuckerman, Jillian York, Robert Faris, and John Palfrey[†]

The Berkman Center for Internet & Society

October 2010

[†] Each co-author is affiliated with The Berkman Center for Internet & Society at Harvard University. Hal Roberts is a fellow; Ethan Zuckerman is a senior researcher; Jillian York is a staff member; Robert Faris is the director of research; and John Palfrey is a faculty co-director of the Berkman Center. We gratefully acknowledge support for this research from the U.S. Department of State via a subgrant through Internews Network (USA).

The Berkman Center shares funders with some of the makers of circumvention tools such as Tor, Psiphon, Freegate and Ultrasurf. Neither the Berkman Center nor the authors of this paper have any involvement in any of the technologies or entities described in this report, including but not limited to Tor, Freegate, or Ultrasurf. Neither the Berkman Center nor the authors stand to profit in any way financially from the success of any general technology or specific tool described in this report.

Circumvention Tool Usage Report

Hal Roberts, Ethan Zuckerman, Rob Faris, John Palfrey

Berkman Center for Internet and Society at Harvard University

Introduction

Circumvention tools allow users to bypass Internet filtering to access content otherwise blocked by governments, workplaces, schools, or even the blocked sites themselves. There are a number of different types of these tools: blocking-resistant tools, simple web proxies, virtual private network (VPN) services, and open HTTP/SOCKS proxies. But every type of circumvention tool provides the same basic functionality – proxying user connections to provide access to otherwise blocked sites. In the following report, we use a variety of methods to evaluate the usage of the first three of these four types of tools to test two hypotheses. First, even though much of the media attention on circumvention tools has been given to a handful of tools – notably Freegate, Ultrasurf, Tor, and Hotspot Shield – we find that these tools represent only a small portion of overall circumvention usage and that the attention paid to these tools has been disproportionate to their usage, especially when compared to the more widely used simple web proxies.[1] Second, even when including the more widely-used simple web proxies, we find that overall usage of circumvention tools is still very small in proportion to the number of Internet users in countries with substantial national Internet filtering.

Key Findings

- We estimate that no more than 3% of Internet users in countries that engage in substantial filtering use circumvention tools. The actual number is likely considerably less.
- Many more users use simple web proxies than use either blocking-resistant tools or VPN services. Of the 11 tools with at least 250,000 unique monthly monthly users, 3 are blocking-resistant tools, 1 is a VPN service, and 7 are simple web proxies.
- When users search for proxy and circumvention related terms in filtering countries, they overwhelmingly search for generic proxy terms like "proxy," and those terms overwhelmingly return either simple web proxies or sites that list simple web proxies and HTTP/SOCKS proxies, not more sophisticated tools.

Circumvention Tool Types

The OpenNet Initiative has documented network filtering of the Internet by national governments in over forty countries worldwide.[2] Countries use this network filtering as one of many methods to

[1] We used google to search nytimes.com (eg. [site:nytimes.com freegate]) for mentions of the ten most popular circumvention tools (as documented in the body of this paper) and found the following number of mentions for the following terms: Freegate: 17, Ultrasurf: 3, Global Internet Freedom Consortium (which includes both Freegate and Ultrasurf): 15, Tor project: 8, Hotspot Shield: 7. The other six tools in the set of ten most used circumvention tools are all simple web proxies – the six combined received only one mention on nytimes.com..

[2] See the country reports, available for free online, at http://opennet.net. For a book-length treatment of this topic, see

control the flow of online content that is objectionable to the filtering governments for social, political, and security reasons. Filtering is particularly appealing to governments as it allows them to control content not published within their national borders. In addition to national Internet filtering by governments, many schools and businesses filter their local connections to the Internet. Many web sites even filter their own content by the geographic location their users – for example, television streaming site hulu.com blocks all users outside of the U.S. from accessing its content.

All circumvention tools use the same basic method to bypass this sort of network filtering: they proxy connections through third party sites that are not filtered themselves. By using this method, a user in China who cannot reach http://falundafa.org directly can instead access a proxy machine like http://superproxy.com/, which can fetch http://falundafa.org for the user. The network filter only sees a connection to the proxy machine (superproxy.com), and so as long as the proxy itself remains unfiltered, the user can visit sites through the proxy that are otherwise blocked by the network filter. Some, but not all, tools also encrypt traffic between the user and proxy, both so that the traffic between the user and proxy is much more difficult to monitor and so, that filtering triggered by the content of the traffic (instead of merely the destination of the traffic) will not work.

Despite this core similarity, circumvention tools differ significantly in many implementation details. We break circumvention tools into four large categories based on their proxy implementations. Each category of tool is distinguished from one another also by virtue of each being closely associated with a single model of financial support. The four categories of tools are:

- blocking-resistant tools
- simple web proxies
- VPN services
- HTTP/SOCKS proxies

The defining characteristic of blocking-resistant tools is that they implement sophisticated methods for evading blocking by filters. A core problem for all circumvention tools is that proxy sites can be blocked just as content and other sites. China can block superproxy.com as well as falundafa.org, and then proxy requests through superproxy.com will cease working. Some tools in each of the above categories use simple forms of blocking resistance to avoid this sort of filtering—for example, a simple web proxy might maintain a list of alternative domain names to send to users in the case that one or more of its existing domain names is blocked. The tools we classify as blocking-resistant tools distinguish themselves from the other categories of tools by implementing much more sophisticated technical means of blocking resistance.[3]

Ronald Deibert, John Palfrey, Rafal Rohozinski, and Jonathan Zittrain, eds., *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace* (MIT Press, 2010).

[3] All of these tools use the same basic method for remaining unblocked. In addition to a core of powerful proxy machines that transit and accelerate proxy traffic, these tools maintain a large pool of "disposable" front end proxies whose only job is to proxy traffic back to the core proxies. These blocking-resistant tools are able to cycle quickly through large numbers of these front end proxies as they are blocked by filtering countries, so that the user is ideally always able to

All of the tools in this category also require installation of a downloaded, client-side application but use the native interface of the browser for web access, unlike the simple web proxies described below. We include only a handful of tools in this category: Freegate, Ultrasurf, and Tor. All of the tools in this category are supported primarily by governmental or charitable funding. Freegate and Ultrasurf are part of the Global Internet Freedom Consortium (GIFC), a Falun Gong-associated group, and are expressly aimed at allowing Chinese people to bypass Chinese government filtering to access political content (though users in other countries use the GIFC tools as well). Tor was initially developed as an anonymity tool but is also effective as a circumvention tool.

Simple web proxies are server-side applications accessed through web page forms. To use one of these tools, the user simply visits a web page that includes a form with a url input box. Instead of entering a web page url into the browser address bar, the user enters the address into the web page form. By submitting this web page form, the user sends the url request to the proxy web server, and the web server returns the page via the proxy. Simple web proxies do not require the user to download or install any client-side application. To use the tool the user needs only visit the web page hosting the proxying web application (for instance http://superproxy.com). Simple web proxies do require, however, that the user navigate the separate, form based browsing interface.

Almost all of these tools support themselves by hosting ads, on either or both of the initial landing page and the banners inserted onto the top of proxied web pages. Some tools go a step further and attempt to replace ads on the requested page with their own ads. Simple web proxies were initially targeted at students in the U.S. and other countries to bypass school filtering systems, but they use the same basic proxying methods as the blocking-resistant tools and so also serve to bypass national filters as long as a given proxy is not specifically blocked by a given country. Some simple proxies have been optimized for usage in countries where the Internet is filtered by the government. At least one widely used proxy site markets itself to users from a specific country, who seek access to YouTube, which is blocked by their government. The only blocking resistance features often used by these tools disguise the proxying functionality of the site, so unlike the blocking-resistant tools, they are most commonly defenseless in the face of IP-based proxy blocking. Some simple proxies register closely related domain names in anticipation of being blocked, and many of their users know that if 1superproxy.com ceases to work in their country, they might try 2superproxy.com.

VPN services use virtual private network software to encrypt and tunnel all Internet traffic through a proxy machine. VPN technology has traditionally been used to allow corporate and other institutional users to access internal networks from the public network, but in the past few years there has been tremendous growth in the availability of personal VPN services. Among other uses, these personal VPN services act as circumvention tools as long as the VPN proxy is hosted outside a filtering country. VPN services might or might not require installation of client-side software (many rely on existing VPN support in Windows or Mac OSX and so need no extra client software) and allow the user to access the web directly through the native browser interface. Because VPN services tunnel all Internet traffic, they can be used for email, chat, and any other Internet service in addition to web browsing. Almost all of these tools support themselves through fees charged directly to users (charges of $10 to $30 per month are common), though a few also offer free services with restricted bandwidth. The

reach some front end proxy that has not yet been blocked. The front end proxies are often provided by volunteer communities, which makes them not only cheaper to operate but more difficult to discover and block since the volunteers have IP addresses randomly distributed across many Internet service providers.

exception to this business model is Hotspot Shield, overwhelmingly the most popular VPN service, which supports itself by injecting ads into the top of all web pages served through its service. VPN services can be blocked either by blocking the IP address they use, or by blocking the protocols used to set up VPN connections. However, since VPNs are often used by business users to access corporate intranets, many governments are reluctant to block the traffic.

The largest class of circumvention tools by number of proxies includes HTTP and SOCKS proxies. These are application level proxies that funnel network traffic through protocols designed to allow web traffic to pass through firewalls.[4] Users generally find lists of these proxies in the form of IP addresses and port numbers on proxy directory web sites. To use a given HTTP or SOCKS proxy, the user enters the IP address and port number of the proxy into a configuration screen of the browser. As a result, no client-side application is needed. The user is able to use the native interface of the browser. These proxies are generally open to the public and have no apparent source of funding (users do not pay to subscribe to them, and the owners of the proxies are anonymous so there is no way to know if they are receiving charitable or government funding).

There are many thousands of these proxies listed on dozens of proxy directory sites, but they turn over very quickly, so the user has to search actively for new proxies to find working ones. With the other three types of tools, it is frequently possible, though often difficult, to discover who is operating a proxy service. With HTTP/SOCKS proxies, it is almost impossible to discover who is running a given proxy, which makes it problematic for many users to place trust in these proxies. Even though it is unlikely that a Chinese government agent or another filtering government is running most of these HTTP/SOCKS proxies, it would be surprising if a Chinese government agent were not running at least some of them given the extreme ease of setting one up. And it is certain that many of these HTTP/SOCKS proxies are run on machines without the knowledge of their owners, for example as botnet proxies.[5]

Research Methodologies and Proxy Usage Estimates

We use three main methods for measuring the usage of circumvention tools – a survey that we conducted, with 134 respondents self-reporting usage numbers on their project or services; analysis of data derived from automated methods regarding web site visits (Google AdPlanner); and analysis of data collected regarded search term frequency (Google Insights). For the blocking-resistant tools and the VPN services, we rely primarily on self-reporting from our survey, since the individual circumvention projects themselves are the only ones with the necessary data about their users. For the simple web proxies, we are able to use Google AdPlanner's data on generic web site visits to measure usage of each tool because each use of a web proxy is also a web page request in itself. We use search frequency as reported by Google Insights as a separate point of reference and confirmation for these other methods. We have no method for measuring the usage of HTTP/SOCKS proxies, since they cannot be measured as generic web site visits and are almost always hosted anonymously.

[4] HTTP is the protocol used by web browsers and web servers to exchange content. HTTP includes in its specification a proxying protocol that allows web browsers and servers to talk to one another through a proxy. The SOCKS protocol is a generic protocol for proxying TCP/IP traffic. Unlike a VPN service, which requires no configuration for individual applications like web browsers, using a SOCKS proxy requires that each application be configured individually to use the proxy.

[5] Threat Matrix. “Botnet proxies hide the bad guy.” <http://threatmetrix.com/resource-center/articles/botnet-proxies-hide-the-bad-guy/>

Measuring usage of circumvention tools is inherently difficult because the purpose of these tools is to obscure the identity of the person who is visiting any given site or sites. The general difficulty of measuring any sort of network usage compounds this difficulty: Is a unique user defined as a unique IP address? A persistent cookie? How can unique users over the course of a day be translated into unique users over the course of a month or a longer period? The problem is further complicated by the need to rely on self reporting for many of the tools, where we have to rely on both the honesty and the competence of each individual tool to measure its own usage. Our goal therefore is not to discover precise levels of usage but instead to find the right orders of magnitude for the usage of individual tools and of classes of tools.

Blocking-Resistant Tool Usage Estimates

We collected the following self reports of usage from Freegate, Ultrasurf, and Tor:

Tool	Unique Users
Freegate + Ultrasurf	500,000 – 1,000,000 monthly
Tor	100,000 – 300,000 daily

The Freegate and Ultrasurf figures represent a combined report for both tools from the developers of the projects based on the combined unique IP addresses connecting to the services over the course of a single month. As with all reports of unique IP addresses, this number likely overstates the number of unique users over a month as many users of DSL and cable modems use dynamically assigned IP addresses that change over the course of a month. For Tor, the usage number is drawn from a 2009 report on usage from the project based on reports from the projects' directory services.[6] Because Tor does not maintain central servers, as Ultrareach and Freegate do, it is not possible to determine how many people are accessing the system – it is possible, however, to see how many unique IP addresses access the directory of Tor entry nodes.

Both of these numbers fluctuate weekly and even daily depending on how aggressively filtering countries, especially China, are blocking these tools. In mid-2010, China was attempting to block these specific tools much more aggressively, and so the usage of each tool was likely nearer the bottom than the top of the estimate range (Chinese users represent a major but not the only source of users for both tools as evidenced by self reports by both projects that successful blocking attempts by China result in significant drops in overall usage). Note that the Tor number is a daily number, and it is unclear how to translate the number directly to monthly users. However, we think it is likely that the two usage estimates are within the same order of magnitude. We make the relatively safe assumption that not every Tor user uses the site daily, and that a monthly usage number is a multiple of the daily usage.

It is possible that there is some overlap in the user base of Tor and Freegate/Ultrasurf tools, as Tor has reported significant interest in their tool from China, and Freegate and Ultrasurf, originally designed for usage in China, now report usage in Iran. That possible overlap suggests usage of blocking-resistant proxy tools may be lower than the monthly and daily users the sites collectively report, but we can

[6] Loesing, Karsten. "Measuring the Tor Network: Evaluation of Client Requests to the Directories." June 25, 2009. <http://metrics.torproject.org/papers/directory-requests-2009-06-25.pdf>

estimate a maximum usage of combined users of the tools by assuming no overlap between them. Assuming no overlap and a very rough conversion of 3 monthly users for each daily user, we estimate the total number of unique monthly users for Freegate/Ultrasurf and Tor as between 800,000 and 1,900,000.

Simple Web Proxy Usage Estimates

To estimate the usage of simple web proxies, we first built a crawler to find as many proxies as possible. The crawler accessed a list of proxy directory sites, recording any links to simple web proxies found on those sites. To include proxy directories and proxies from a range of different countries, we seeded the crawler with the top ten proxy-related Google search results of the most common proxy related search term (as described below in the *Search Frequency* section) for each of the following countries: China, Russia, Iran, Egypt, Tunisia, Burma, Vietnam, Kazakhstan, Uzbekistan.[7] For each search result, we either included the proxy directly if the result was itself a simple web proxy or included all of the proxies listed by the directory if the result was a proxy directory.

We ran the crawler daily for three months from February to May of 2010, collecting any new proxies on the identified proxy directories over that period. During this period, we found a total of 11,350 simple web proxies. We think this significantly undercounts the total number of simple web proxies advertised during this period, since we have no reason to believe that our list of proxy directories was comprehensive or that every simple web proxy is listed in a proxy directory.[8] However, we assume that the most popular simple web proxies will be advertised across several proxy directories, so we are less likely to miss the biggest proxies.

We took each of those proxies and retrieved the unique monthly users from Google AdPlanner for each. Google AdPlanner uses a variety of data sources to estimate unique visitors to any web site domain, including google.com searches, Google toolbar data, Google Analytics data, consumer panels, and third party market research data.[9] Because each proxied request through a simple web proxy is itself a request for a web page, these web traffic data accurately reflect the usage of simple web proxies to the degree that the web traffic estimates are accurate. These AdPlanner estimates are based on Google's vast trove of data about web traffic, and so we view them to be as accurate as any third party estimate of web traffic.

The data from Google that we use to measure usage is the AdPlanner estimate of unique monthly users for each web site. Google uses unique cookies as the basis for this user estimate, but it adjusts the unique cookie number down to recognize the fact that cookies do not directly correspond to users (the same user often uses multiple computers with different cookies, for instance). Any cookie based estimate is likely to be smaller than an IP-address-based estimate of users, since IP addresses change more frequently than cookies. Thus, AdPlanner's estimate of monthly users is less than a cookie-based

[7] This list of countries was chosen to be a geographically diverse set of countries that implement some level of filtering on their national networks.

[8] We did not include http://proxy.org in our crawling because they specifically requested that we not crawl their site. Proxy.org alone lists up to 9,000 simple web proxies a day in its service, an indication that we have undercounted the total number of simple web proxies.

[9] Google. “How Doubleclick Ad Planner data is generated.” 2010. <http://www.google.com/support/adplanner/bin/answer.py?answer=98132>

estimate, which is likely to be less than an IP address-based estimate.

Of the 11,350 proxies found by the crawler, AdPlanner returned web traffic data for only 183 of them. We believe this is because only 183 proxies meet Google's minimum requirements for “significant traffic.” Google does not disclose its minimum threshold for “significant traffic,” but the lowest number of unique users returned for any site we submitted to AdPlanner was 3300 users. We therefore assume that the threshold is around 3300 users. After the three month crawler run completed, we manually tested 50 proxy urls chosen randomly from the proxies for which AdPlanner returned no data. Of the 50 tested proxies, 40 were no longer valid proxies. This is not an unexpected result, since proxy operators frequently churn through proxy urls as proxies get blocked, become unprofitable, lose interest for the operator, and so on. We assume that the remainder of the estimated 20% of proxies that were working proxies but returned no data from adwords had less than 3300 users per month.

From the 183 proxies for which AdPlanner returned data, the ten with the most unique users are below. We have replaced the names of the proxies with numbers to avoid creating a roadmap for organizations that would filter these proxies.

Simple Web Proxy	Unique Monthly Users
SWP 1	1,100,000
SWP 2	840,000
SWP 3	760,000
SWP 4	520,000
SWP 5	430,000
SWP 6	390,000
SWP 7	270,000
SWP 8	220,000
SWP 9	180,000
SWP 10	150,000

Including the proxies above, we found 15 simple web proxies with at least 100,000 unique monthly users. The average number of unique monthly users for the 183 proxies was 49,620, and the median was 14,000. The total of all users from all 183 proxies is 9,080,400, but that total over counts the total number of estimated unique users of simple web proxies because the users are only unique for each proxy (the same user may use a number of different simple web proxies, and we have no way of measuring the degree to which proxies share users).

In aggregate, our method suggests an estimated user base for simple web proxies of between 10-15 million. In aggregate, we believe usage of simple web proxies is at least an order of magnitude larger than use of blocking-resistant proxy tools but is still a very small portion of all Internet users.

VPN Service Usage Estimates

To estimate the usage of VPN services, we spent over a hundred hours searching the Internet for advertised VPN services and then surveyed each of the services for their daily and monthly usage numbers. We also recorded various characteristics of each VPN, including the date of creation and the funding model. In total, we found 134 VPN services. Even though we took considerable effort to exhaust the search, it is certain that we missed some services. In particular, the search for VPN services was done by an English speaking researcher and so focused on English language advertised services (even though we found a number of foreign language-based services through the search). It is possible that we missed a significant number of foreign language based services through this method.

Of those 134 VPN services, 90 have been created in the past three years according to their domain creation dates. Almost all of the services charge a subscription fee. Hotspot Shield, by far the most popular VPN service we found, is the exception to this rule— it is free and supports itself by injecting ads into all web pages retrieved over its network. We found three other free VPN services: one that is ad supported like Hotspot Shield (but the subject of many reports of poor reliability and even spam campaigns aimed at its users), one that claims to be in a beta period for testing, and one that is run by a company that sells filtering software.[10] We found five other services that offer a free version with limited bandwidth as an alternative to their subscription-based service. Many sites offer free services that are hobbled in more serious ways, for example, many are either limited to a trial duration of a few days or requiring reconnection to the network every five to twenty minutes.

We sent an email survey to each of the 134 services and received responses from 21, a 15.7% response rate. Many of the responses returned usage data in a form other than daily or monthly unique users, requiring us to use very crude conversions to estimate the order of magnitude of the usage for each tool. Daily usage was the statistic most commonly reported, so we applied a rough conversion from monthly unique users to daily unique users for sites that only reported only monthly users. Several sites reported both daily and monthly usage, which allowed us to establish a 3:1 ratio for monthly:daily traffic statistics. Some respondents reported non-unique daily users. Based on respondents who reported both numbers, we assumed a ratio of 6 non-unique daily users to 1 unique daily user. Some respondents reported only concurrent users. Again, extrapolating from other responses, we assumed a ration of 1 concurrent user to 4 daily unique users.

Hotspot Shield did not respond to this survey but previously reported their usage as 600,000 unique users per month. We do not include Hotspot Shield in the aggregate survey results below because we consider them to be an outlier among VPN services, both because of their usage and because of their unique status as a viable, reliable, free VPN service.

Based on the self reports from the survey and on the rough conversions described above, we estimate that the 22 responding VPN services had a mean of 3909 unique daily users and a median of 1667 unique daily users. Of the 22 respondents, only 3 had at least 10,000 users (their numbers were 10,500; 12,300; and 30,000). If we add together the unique users of each tool, we find only 82,089 unique daily users among the responding VPN services, excluding Hotspot Shield. If we assume that the responding VPN services are representative of the whole pool of 134 services, the 134 services would total approximately 500,000 unique daily users. Given the cost associated with subscribing to a VPN service, it is unlikely there is much overlap between users of VPN services, as we expect to find among

[10] We leave as an exercise for the reader deciphering the considerable value for a filtering company of running its own circumvention system.

users of simple web proxies.

It is possible that the 21 responding VPN services are significantly smaller than the population of 134 VPN services (and of the larger population of all existing VPN services) or that VPN services with very large user bases were not among the respondents. To test whether the 21 responding VPN services under-represent large VPN services, we compared the Google search frequency for the domain of each of the 134 VPN services with the search frequency for the term “hotspotshield” (see below for further details about the search frequency method). So, for instance, for a hypothetical VPN service with a domain name of “supervpn.com,” we compared the frequency of searches for “hotspotshield” to those for “supervpn.” We found only 11 VPN services whose name had a search frequency of at least 1/80 the size of search frequency for “hotspotshield.” Of those 11, only two had a search frequency greater than 1/80: the free VPN service run by a filtering company had a search frequency of 2/80, and a VPN service sold by a company that also sells other more popular privacy software had a search frequency of 3/80. Of the remaining 9 VPN services that registered a search frequency of 1/80, 4 were represented among the respondents of our survey. This over-representation of respondents among the group of the most searched for VPN services provides some assurance that our survey responses do not under-represent large VPN services. And the finding that the VPN service with the greatest search frequency other than Hotspot Shield had only 3/80 of the searches of Hotspot Shield is evidence that we have not missed any very large VPN services among the 134 services we found.

Combining Hotspot Shield's reported numbers and our estimate above, assuming no overlap, and using the rough 1:3 conversion of daily to monthly users, we roughly estimate 2.1 million users of VPN services, the same order of magnitude as users of blocking-resistant proxy servers.

HTTP/SOCKS Proxy Usage Estimates

There are at least tens of thousands of open HTTP/SOCKS proxies advertised on various proxy directories, based on our incidental collections of such proxies through the crawler described above. It is common for a single proxy directory to list thousands of such proxies at any given time. But there is no reasonable way to reliably determine the usage of such tools. The only way to directly measure usage is on the proxies themselves, and the vast majority of the proxies lack any contact information that could be used to survey usage. One method of measuring the usage of the proxies would be to analyze the logs of a filtered websites from various countries and try to identify HTTP / SOCKS proxies among the connecting IP addresses, but we did not attempt this method because of the difficulty of obtaining a range of such highly sensitive logs and the difficulty of identifying HTTP / SOCKS proxy IP addresses with sufficient reliability.

It is worth noting that HTTP/SOCKS proxies require more user knowledge and involvement than simple proxies. A user needs to edit her network settings to specify a proxy, and may need to change proxy settings frequently. Users in cybercafés and some other settings are often prevented from accessing a machine's network settings. Given these factors, we suspect HTTP/SOCKS proxies are used less often than simple proxies, but it would be very difficult to definitively confirm that suspicion.

Search Frequency

Google Insights provides data on the relative frequency of search terms globally and within specific countries. An alternative method for testing the popularity of various circumvention tools and methods is to use Google Insights to determine the search frequency for related terms. We tested the popularity of twenty circumvention related search terms in nine different countries (China, Russia, Iran, Egypt, Tunisia, Burma, Vietnam, Kazakhstan, and Uzbekistan), including both the English terms and the local language translations of each term. We also collected euphemisms for filtering circumvention in each local language from local experts in each country and tested the popularity of those terms. For example, experts reported to us that "climbing over the wall" is commonly used to search for tools to bypass the "great fire wall" in China. The tested terms included some of the most popular tools from each of the blocking-resistant, simple proxy, and VPN service tool types.

We found that in every country except China and Iran, the most popular proxy-related search term by a large margin was the English language term "proxy." In China, the most popular search term was the direct Chinese translation of "proxy." In Iran, the most popular search term was the Farsi translation of "filter." We combined the first ten search results from the most popular single search term in each of the nine countries. This combined set consisted of 49 proxy directories, 33 sites unrelated to circumvention, 8 proxy listing blogs, 8 simple web proxies, and 2 redirectors to simple web proxies. The proxy directories overwhelmingly listed a combination of simple web proxies and HTTP/SOCKS proxies. None of the results directly pointed to blocking-resistant tools or to VPN services.

These results are not definitive. Google searches are one of several discovery methods for circumvention tools, and they are also a reflection of general cultural awareness of a given term. Certainly in some countries Google searches for specific circumvention related terms are blocked or edited, and even where Google search results are not directly edited the sites of many tools are blocked, skewing the search results away from those blocked sites (and presumably decreasing the popularity of the fruitless searches). But we think that these results provide further evidence for the finding that usage of simple web proxies is at least as large as, and probably considerably larger than, usage of the blocking-resistant tools.

Conclusions

Usage estimates for blocking-resistant tools and for simple web proxies suggest that simple web proxies are at least as popular as the blocking-resistant tools and are likely an order of magnitude more popular, in aggregate. Of the 11 circumvention tools with at least 250,000 monthly users (Ultrasurf, Freegate, Tor, Hotspot Shield, and SWP #s 1 – 7), 7 are simple web proxies. Those 7 proxies together appear to serve close to half of the combined unique users of the 183 simple web proxies whose usage we were able to estimate. The number of subscription-based VPN services has more than tripled over the past three years, but the usage of these services, other than Hotspot Shield, is still a relatively small portion of circumvention tool users, totaling about as many users as the largest single blocking-resistant tool or simple web proxy. Search frequency data further shows that searches for circumvention related topics is dominated by forms of the generic term "proxy" and that those searches overwhelmingly return simple web proxies or proxy directories dominated by simple web proxies and HTTP/SOCKS proxies.

This result should not be interpreted to diminish the importance of blocking-resistant tools or VPN services. Tor provides an important anonymizing service as well as enabling circumvention of filtering, and Freegate, Ultrasurf and VPN systems allow users in nations that aggressively filter the Internet to obtain relatively uninterrupted connections to the Internet. VPN services provide significantly more functionality than simple web proxies because they proxy the entire network connection. But this result does suggest that scholars, advocates, and others need to take seriously the role simple web proxies play in enabling circumvention of Internet filtering.

We were surprised to discover that several widely-used simple proxies remained unblocked for very long periods of time in highly censorious nations that aggressively block the more well-discussed blocking-resistant tools. This difference in the treatment of the different types of tools may be the result of the difference in press coverage of these tools. Unlike Freegate, Ultrasurf, and Tor, the more widely-used simple web proxies have not been lauded much if at all in the U.S. press as agents of political change. One explanation for why the more widely used but less widely discussed tools have not been blocked as aggressively is that decisions about which tools to block are based more on political considerations (including both the politics of the projects themselves and the political slant of media coverage of the projects) than on technical considerations. This hypothesis presents a difficult challenge for those hoping to support circumvention tools as agents of political chance, in particular for the U.S government, because any support by organizations or the U.S. government for these tools may have the side effect of increasing the efforts of filtering governments to block the supported tools.

It would be unwise to assume that all the users we estimate are using simple proxies are located in nations that aggressively filter the Internet. We know from interviews with operators of ad-supported proxies that they target their sites towards students circumventing school firewalls and other users in nations with little or no national filtering. Field studies suggest that ad-supported proxies are often used to evade server-side geographic restrictions on otherwise uncontroversial content in countries that do not filter the Internet. Users of blocking-resistant tools, on the other hand, are more likely to be located in countries with substantial filtering.

Altogether, we find that, even given the large margin of error for our estimates, usage of all of the tools described here is very small compared to the total population of approximately 2 billion Internet users globally or even the population of users in countries that aggressively filter the Internet. The OpenNet Initiative lists 13 countries as implementing "substantial" Internet filtering, blocking content for political, social or security reasons.[11] The estimated number of Internet users in those thirteen countries totals 562 million, and we would expect many of those users to turn to circumvention tools to access blocked parts of the Internet.[12] The top end of our estimates for blocking-resistant proxies, simple proxies and VPN services totals 19 million (1.9 million for blocking-resistant tools + 15 million for simple web proxies + 2.1 million for VPN services).

If we assume no overlap between those tools and assume that they are only being used by citizens of the nations where filtering is most pervasive, only 3% of Internet users in these countries would be using blocking-resistant proxies, simple proxies, or VPN services. We do not include HTTP/SOCKS proxies in that 3%, but the usage of HTTP/SOCKS proxies would have to be much higher than that of the other classes of tools to significantly raise the usage of circumvention tools, even including the various assumptions that overestimate usage, including especially the large number of users of these

[11] Deibert et al., *supra* note 2.

[12] Internet usage statistics according to International Telecommunications Union, http://www.itu.int/ITU-D/icteye/Reporting/ShowReportFrame.aspx?ReportName=/WTI/InformationTechnologyPublic&ReportFormat=HTML4.0&RP_intYear=2009&RP_intLanguageID=1&RP_bitLiveData=False [last accessed on September 22, 2010].

tools in countries without “substantial” filtering.

The relatively small usage of circumvention tools, even in filtering countries, suggests either that users do not know that the tools exist, do not know how to find them, or consider that the benefits of using the tools do not outweigh the costs for most users. The relative popularity of “proxy” searches documented in this paper indicates that at least some users are able to find them with simple search terms and that those search terms are unblocked (otherwise they would not be popular), but the high popularity of technical terms like “proxy” relative to other proxy related search terms might also mean that only technical users are aware of the tools. Even if users are able to find the tools, it may be that some combination of the usability, performance, and security of the tools is not good enough that users find the benefit of circumventing filtering worth the cost of using the tools. Even though we have found the usability of most of the tools to be adequate, any effort required to install and use the tools will dissuade some number of potential users. Performance is likely to be a more significant problem. Performance tests from a 2007 evaluation of circumvention tools by some of the authors of this paper suggest that performance was a significant problem for all circumvention tools in 2007, and anecdotal experience suggests that it continues to be a problem.[13] The security properties of the tools has a very unclear effect on usage, since we have neither a sense of users' perceptions of the security risks of using the tools nor of users' perceptions of the security properties of particular tools. Finally, it may be that there is just not as much interest in circumventing Internet filtering as widely believed for any of a number of reasons. For example, users in many filtering countries may simply prefer to access local content, written in their own languages about topics of local interest, despite the fact that the local content is subject to traditional government regulation and therefore highly censored.[14] We note that three of the nations that have at least tens of millions of Internet users and who aggressively filter the Internet – China, Iran and Vietnam – have made significant investments in creating locally hosted alternatives to popular social media platforms like YouTube and Facebook. Our findings may suggest the logic of this approach – a large percentage of users in nations that aggressively filter the Internet either do not know how to conveniently reach these popular sites, or they have decided to use censored, local alternatives.

[13] Roberts, Hal et al. *2007 Circumvention Landscape Report: Methods, Uses, and Tools.* Berkman Center for Internet & Society. 2009. <http://cyber.law.harvard.edu/publications/2009/2007_Circumvention_Landscape_Report>

[14] For example, pending research by the authors of this paper shows that about 95% of web page visits in China are to web sites hosted within China, which may show that filtering has been extremely successful or simply that Chinese people like to read web sites written in Chinese by other Chinese people about topics of local interest to China.