Lost in Transmission: Investigating Filtering of COVID-19 Websites

Anjali Vyas, Ram Sundara Raman, Nick Ceccio,
Philipp M. Lutscher*, Roya Ensafi

University of Michigan, *University of Oslo

Abstract. After the unprecedented arrival of the COVID-19 pandemic, the Internet has become a crucial source of essential information on the virus. To prevent the spread of misinformation and panic, many authorities have resorted to exercising higher control over Internet resources. Although there is anecdotal evidence that websites containing information about the pandemic are blocked in specific countries, the global extent of these censorship efforts is unknown. In this work, we perform the first global censorship measurement study of websites obtained from search engine queries on COVID-19 information in more than 180 countries. Using two remote censorship measurement techniques, Satellite and Quack, we collect more than 67 million measurements on the DNS and Application layer blocking of 1,291 domains containing COVID-19 information from 49,245 vantage points in 5,081 ASes. Analyzing global patterns, we find that blocking of these COVID-19 websites is relatively low—on average, 0.20%-0.34% of websites containing information about the pandemic experience interference. As expected, we see higher blocking in countries known for censorship such as Iran, China, and Kazakhstan. Surprisingly, however, we also find significant blocking of websites containing information about the pandemic in countries generally considered as "free" in the Internet space, such as Switzerland (DNS), Croatia (DNS), and Canada (Application layer). We discover that network filters in these countries flag many websites related to COVID-19 as phishing or malicious and hence restrict access to them. However, our investigation suggests that this categorization may be incorrect—most websites do not contain serious security threats—causing unnecessary blocking. We advocate for stricter auditing of filtering policies worldwide to help prevent the loss of access to relevant information.

Keywords: Censorship · COVID-19 · Filtering · Phishing

1 Introduction

The COVID-19 pandemic has necessitated heavy reliance on the Internet by people all over the world. Essential information about the pandemic, including details about the virus and the disease, state- and country-level spread, guidelines, and tracing are primarily accessed through the Internet [21]. However, at the same time, there has also been a surge of misinformation which has prompted

authorities to exercise greater control over Internet resources [34]. Although restricting access to malicious resources may be necessary in order to protect users, several studies have shown that access to legitimate information also may be restricted [12, 24, 38, 55].

Recent work by the censorship measurement community has pointed to the blocking of specific websites related to the COVID-19 pandemic in certain countries. OONI [45], a censorship measurement platform, investigated Myanmar's government directive that all Internet service providers must block websites supposedly containing "fake news" regarding the pandemic [24]. In addition, the Citizen Lab [46] found that sources of COVID-19 information that criticize the government are being actively censored on Chinese social media [12, 38]. These investigations reveal that governments are engaging in possibly detrimental censorship of COVID-19 information. However, efforts to measure censorship of COVID-19 information have so far been restricted to certain countries and a small number of websites. The global extent of blocking of legitimate COVID-19 information is as yet unknown.

In this paper, we present the first global censorship study of websites that provide potentially factual information about the pandemic. We use two recently introduced remote censorship measurement techniques, Satellite/Iris (we use just "Satellite" for briefness) [32, 39] and Quack [48] and study the DNS and Application layer blocking (respectively) of 1,291 domains related to COVID-19 in more than 180 countries. Specifically, we aim to answer the following research questions:

1. What is the share of COVID-related websites blocked?
2. Where are COVID-related websites blocked?
3. What categories of COVID-related websites are blocked?
4. Why are COVID-related websites blocked?

To answer these research questions, we first gather a list of 81 neutral search terms that yield potentially factual information about the pandemic from Google Trends [21]. We then perform geo-distributed search engine crawls using three popular search engines in nine different countries. We collect the top ten websites from each crawl, resulting in a set of 1,291 domains most related to the pandemic (which we refer to as "COVID-related test list"). We then perform remote censorship measurements to 29,113 Satellite vantage points and 20,989 Quack vantage points, resulting in a pool of 67 million measurement points. We make our list of domains and measurement data public for other researchers to use [49]. We additionally add over 86 million measurements for domains that are popular [2] and politically sensitive [10], but not strictly related to the pandemic. This set of domains (which we term "Censorship Measurement test list") has been used extensively by censorship studies in the past [43, 45] and provides a point of comparison.

Analyzing patterns in the data, we find that the global blocking of websites in the COVID-related test list is relatively low—On average, 0.20%-0.34% of the websites on our COVID-related test list experienced interference compared to 0.70%-1.04% of websites in the Censorship measurement test list. As expected,

we see more blocking in both test lists in countries known for censorship such as Iran, China, and Kazakhstan. However, more surprisingly, our measurements show significant blocking of COVID-related websites in many countries with high Internet freedom scores [11] such as Switzerland (DNS), Croatia (DNS), and Canada (Application layer). Upon investigation, we find that networks in these countries employ web filters such as Fortiguard [19], which categorize many COVID-related websites as containing phishing or other malicious content, resulting in their unavailability from vantage points in these networks.

We utilize different URL classifying services [7, 29, 50] and manual investigation to determine whether 46 COVID-related websites blocked by web filters have harmful phishing or other malicious content. Interestingly, while Fortiguard classifies 91.30% of the websites as phishing or malicious, our results show that only 0-36.96% of websites are marked as containing security risks by other services, illustrating the wide variance in categorization and, transitively, blocking policies of web filters. Our manual investigation further suggests that only 2.17% of websites actually contain harmful and evident security threats.

Our findings show that such 'benevolent blocking' may restrict the amount of factual and, in some cases, essential information available on the Internet. With that in mind, we advocate for stricter auditing of censorship policies and for more transparency regarding what is being blocked by groups making use of these filtering services. Only with such transparency can we ensure that valuable information is kept open and available to Internet users.

2 Background and Related Work

The COVID-19 Pandemic and the Internet On March 12, 2020, the World Health Organization (WHO) officially declared the Coronavirus outbreak as a pandemic. At the time of writing, more than two million deaths were confirmed caused by COVID-19 worldwide [53]. Whereas there has been, and there still is, a lot of variation in how governments respond to the pandemic, most governments imposed regional or country-wide shutdowns, banned mass gatherings, encouraged social distancing, made the wearing of face masks mandatory, and invested in their health care systems to slow down the spread of the virus [8].

Apart from direct disease control, many authorities also increased their efforts in controlling (mis)information during this global pandemic [3]. Reports emphasize that conspiracy theories and disinformation attempts related to COVID-19 have drastically increased [28]. Several studies show that bots and ordinary users promote misinformation on social media [6,9,18,41]. As a response, many authorities enacted policies to counter this so-called "infodemic." Governments passed laws to criminalize falsehood related to public health, created special units to remove disinformation, and delegated this task to social media or private Internet companies [34, 55]. However, there is tentative evidence that legitimate information on the pandemic is also blocked. For instance, reports by the Citizen Lab and the New York Times show that regime-criticizing information on the pandemic is actively censored in Chinese social media [12,38,55]. A report by OONI

shows that the Myanmar government has been ordering long Internet shutdowns and blocking COVID-19 related content in a non-transparent manner [24]. The pandemic offers an opportunity to intensify online censorship efforts and undemocratic policies [22, 25]. However, these case studies only highlight specific cases of censorship in a few countries. To our knowledge, our study is the first that systematically investigates the share of online blocking of COVID-19 related websites worldwide.

Censorship Studies Censorship mechanisms vary across countries and networks, and therefore many censorship measurement techniques have been proposed to measure and quantify what is being blocked and how the blocking is occurring. On a technical level, network censorship is defined as the deliberate disruption or blocking of certain types of Internet communication by a network adversary. At the coarsest level, an adversary may prevent access to Internet connectivity completely for a user population, a phenomenon termed as *Internet shutdowns* [13,14]. These are out of the scope of our study. Rather, we investigate *Internet censorship* where access to specific websites is blocked. There are commonly three stages of a network connection that could be blocked. First, a censor may restrict access during the TCP handshake stage of an Internet connection between a client and server, based on the server's IP address. This method is not widely used because of the emergence of Content Delivery Networks (CDNs), but is still used to block access to circumvention proxies [1]. Second, a censor may inject a DNS query response with a non-routable IP, an IP that leads to a blockpage, or may not return an IP at all [4, 32]. Finally, a censor may also inspect specific HTTP and TLS packets and on observing a particular keyword, reset the connection, inject blockpages or drop packets [44,48]. In this paper, we focus on DNS poisoning and application-layer blocking as they are two common methods of censorship implementation.

Censorship measurements can be conducted from within countries of Interest ("Direct Measurement") or remotely from outside the country ("Remote Measurement"). *Direct Measurement* uses volunteer devices or accessible vantage points inside countries to send network packets to possibly blocked hosts. There has been a plethora of studies that have directly measured censorship within a specific country [5,17,26,45,51,54,56]. This kind of measurement is highly useful for in-depth analysis of censorship, but due to scale, coverage, continuity, and safety limitations is not ideal for widespread global measurement [43].

More recently, *Remote Measurement* techniques that can measure censorship without accessible vantage points or volunteers have enabled global measurements of high scale and coverage [31,32,39,42–44,48]. These techniques use side channels in existing Internet protocols for interacting with remote systems, and infer whether the connection is disrupted from their responses. In this paper, we use two types of remote measurement techniques, Quack and Satellite.

– **Satellite** Satellite sends DNS requests from a single measurement machine towards many infrastructural Open DNS resolvers and control resolvers in different countries [32, 39]. Satellite then compares the responses from the

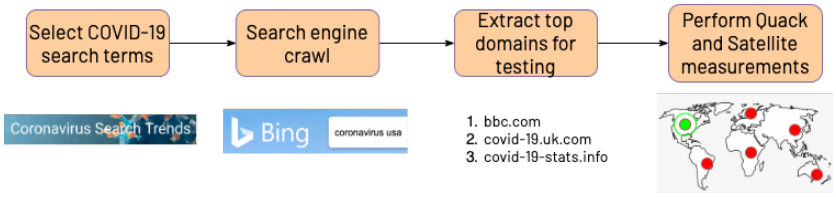

Fig. 1. Flowchart of methodology steps.

Open DNS resolvers and control resolvers using a set of 5 heuristics to determine the presence of network interference [32].

– **Quack** Quack uses infrastructural servers that have the TCP Echo functionality enabled on Port 7 as vantage points to measure censorship of specific keywords [48]. Quack uses a retry-based mechanism to send HTTP-lookalike requests containing both sensitive and benign payloads to the Echo server vantage point. In the absence of any censorship, both types of requests would be reflected back to the sender as is. However, in case the sensitive keyword is censored (through injecting a blockpage, reset, or forcing the connection to timeout), the expected response would not be received. Quack also uses Echo's sibling protocol, Discard, to determine directionality of blocking. In the case of Discard, the remote vantage point is expected to drop all of the packets, but a censor acting on incoming packets may choose to inject a reset or a blockpage.

3 Methodology

To collect data on the blocking of potentially important information related to the COVID-19 pandemic, we assemble a list of search keywords that yield factual information related to COVID-19 and perform search engine crawls to gather popular domains. We then test reachability to these domains using remote censorship measurement techniques. Figure 1 provides a flowchart summarizing the data collection methodology.

Selection of search engine crawl keywords We first use Google Trends to assemble a list of 81 different search terms meant to yield factual information on COVID-19 in search engine results [21]. Google Trends provides data about the most common Internet searches related to the pandemic performed by Google Search users. We note that most search terms provided by Google consisted of the word 'coronavirus' followed by another word or the name of a country, such as 'coronavirus cases' or 'coronavirus usa.' We add 26 such keywords to our list, followed by the same keywords with the word 'coronavirus' replaced with 'covid' or 'covid-19.' Finally, four general terms ('coronavirus', 'corona virus', 'covid', and 'covid-19') complete the list. We utilize this list of frequently-searched keywords as they are more likely to yield factual and important information on COVID-19 that should be available to anyone around the world.

Table 1. Distribution of vantage points used for measurement (CR: Covid-related test list, CM: Censorship Measurement test list).

Technique	# VPs		# Countries		# Autonomous Systems (ASes)		Median # of ASes Per Country	
	CR	CM	CR	CM	CR	CM	CR	CM
Satellite	29,113	28,415	165	166	4,073	3,920	5	5
Quack Echo	20,799	10,607	151	125	2,089	1,350	3.5	3
Quack Discard	7,730	7,993	112	112	1,165	1,184	3	3

Forming the Test List Using the list of 81 search terms, we perform search engine crawls to gather the URLs of websites containing information on COVID-19. To ensure that our list of URLs accurately reflect genuine websites people across the globe would access for COVID-19 information, we execute this crawl on nine different geo-distributed vantage points located in England, France, Germany, Ireland, Canada, Japan, South Korea, Singapore, and Australia. Using Selenium [40], we query Google, Bing, and DuckDuckGo with each of our search terms, recording the URLs of the top ten websites.

We take the union of the list of URLs recorded, which results in a list of 4,155 unique URLs hosted on 1,291 live domains. We use these 1,291 domains, termed as the *COVID-related test list* as input to our measurements testing for blocking. Since these websites form top search results for the different countries, a censor aiming to block factual information on COVID-19 would likely block these websites. These websites fall into 43 categories according to categorization by Fortiguard's URL filter service [19]. The most common categories are News and Media, Government, and Health.

In addition to these 1,291 COVID-related domains, we also create an additional *Censorship Measurement test list* composed of 2,128 sensitive and popular domains from Citizen Lab [10] and Alexa [2] that are regularly tested by other censorship measurement platforms [43, 45]. The overlap between the COVID-related test list and the Censorship Measurement test list is very small, consisting of only 70 domains, and as such, the two test lists provide a point of comparison.

Censorship Measurement We use Quack and Satellite to determine whether the domains in the test input lists are being filtered. Measurements using these techniques were performed for the COVID-related test list and the Censorship Measurement test list over a period of two weeks, from June 12, 2020, to June 26, 2020, from different machines in North America. For Quack, we performed both Echo and Discard measurements. The number and distribution of vantage points used by each technique for measurements (of the COVID-related and Censorship Measurement test lists) is shown in Table 1.

Ethics We follow all the recommendations made in previous studies that have performed remote censorship measurements [31, 32, 35, 43, 44, 48] and have only used "infrastructural" vantage points. Specifically, we only use nameservers for

DNS measurements [32] and servers and routers for Quack measurements in countries with strict Internet control [48]. We also follow all the Internet measurement recommendations made in the line of work using Internet-wide scans such as ZMap [16]. We rate limit our measurements, close all connections, and host a web server on our measurement machines which provides details of our research and offers administrators the option to opt-out.

Data Analysis Overall, we collect around 153 million censorship measurements using our list of vantage points and the two test lists. We perform around 67 million measurements for our COVID-related test list. We augment our measurements with country information from Maxmind [27] and AS information combined from Maxmind [27], Routeviews [37], and Censys [15]. We perform measurements in 186 countries and 5,081 Autonomous Systems (ASes).

Our measurement techniques perform multiple probes during each test, and the test is marked as interfered only if all the probes fail. This helps to prevent false positives from momentary glitches in the network. In addition, we manually remove false positives originating from rogue vantage point responses and use blockpage and false positive fingerprints recorded in previous studies [43, 44] to label our data and avoid false inferences.

We next calculate the average blocking rate across each of the countries covered by our measurements. More precisely, we calculated the average blocking rate in a country cc with n vantage points as:

$$\text{Avg. Blocking Rate}_{\text{cc}} = \frac{\sum_{i=1}^{n} \%\ \text{domains blocked}_{\text{vp}_i}}{n} \tag{1}$$

We use this quantitative value in our results. For our country-level aggregates to be more accurate, we only report aggregate results for countries with 10 or more vantage points in our results.

4 Results

The worldwide measurement of COVID-19-related websites allows us to answer our research questions outlined in the introduction.

4.1 What is the share of COVID-related websites blocked?

On a positive note, the global average blocking rate of COVID-19 related websites seems to be relatively low. On average, only 0.20%-0.34% (depending on the protocol tested) of websites experience some sort of interference. This is lower compared to an average blocking rate of 0.70%-1.04% per country from the Censorship Measurement test list of politically sensitive and popular domains. Nevertheless, our measurements still find many COVID-related websites filtered in networks in a considerable number of countries. Perhaps the most surprising finding is that several countries previously not known for Internet censorship observe the highest blocking rates for these websites.

Table 2. Top five countries having the highest average blocking rate across the three sets of domains(CC: Covid-containing, CR: Covid-related, CM: Censorship measurement) in Satellite, Quack Echo, and Quack Discard.

Satellite			Quack Echo			Quack Discard		
CC	*CR*	*CM*	*CC*	*CR*	*CM*	*CC*	*CR*	*CM*
CH (4.32%)	CN (10.74%)	CN (15.71%)	EC (2.50%)	IR (8.98%)	IR (29.50%)	CN (1.52%)	IR (7.77%)	IR (33.27%)
HR (2.39%)	IR (1.76%)	IR (14.95%)	CN (1.17%)	CN (4.30%)	CN (11.81%)	CA (1.42%)	CN (4.45%)	CN (11.44%)
KZ (2.23%)	KZ (0.57%)	IQ (2.96%)	IR (1.09%)	EC (2.29%)	BD (2.94%)	TW (0.78%)	VN (0.37%)	KZ (1.82%)
AU (1.55%)	SG (0.56%)	ID (2.46%)	CA (0.82%)	SI (1.25%)	PK (2.48%)	IR (0.48%)	EG (0.28%)	TR (1.57%)
DK (1.26%)	CH (0.52%)	AF (2.10%)	BD (0.79%)	TN (0.71%)	TN (1.74%)	RO (0.31%)	RU (0.19%)	EG (0.93%)

To showcase this, we create an additional set of domains from the COVID-related test set that consists exclusively of the domains that have the phrases "covid", "corona" or "korona" in them. These domains likely became live after the pandemic started with the purpose to provide users with information related to COVID-19. We call this list *COVID-containing* and use it as an indicator of blocking specifically related to the pandemic. There are 1,291 distinct domains in our COVID-related set, out of which 152 are in our COVID-containing set. These websites appear in the top search engine results for common COVID-19 queries, and as such may provide useful information to Internet users on the pandemic.

Table 2 shows the top 5 countries in which we observe the highest average blocking rates for these three sets of domains in Satellite and Quack. Whereas we observe the highest average blocking rate in China and Iran in most test lists, countries previously not known for Internet censorship (Switzerland, Croatia, and Canada) appear in the top 5 in the COVID-containing and COVID-related test lists.

To investigate this finding more systematically, we correlate our measurements to a qualitative Internet censorship measure quantified by "Varieties of Democracies" [11]. This measure is judged for 202 countries by several country experts. Figure 2 illustrates the results. The labeled countries exhibit blocking that is higher than 90% of blocking observed in all countries. A simple linear regression shows a positive correlation between the level of Internet censorship and the average blocking rate in most test sets. Nevertheless, in particular for the COVID-containing list, we find many countries with low censorship scores from "Varieties of Democracies" that experience relatively high website blocking rates in our tests.

4.2 Where are COVID-related websites blocked?

Based on the results in Table 2 and Figure 2, we analyze the blocking of COVID-related domains for certain countries in detail. We first explore two countries,

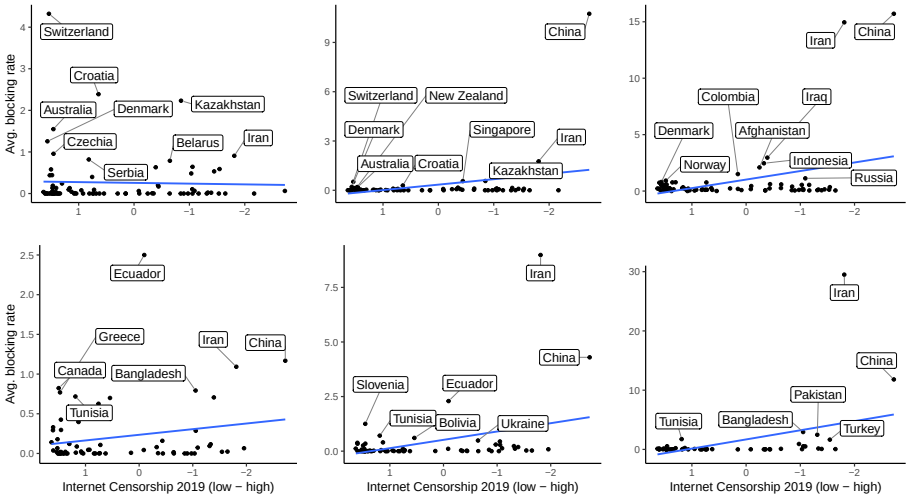

Fig. 2. Correlation between qualitatively measured Internet censorship level (2019) and blocking with Satellite (top) and Quack Echo (bottom) for the *Covid-containing* (left), *Covid-related* (middle) and *Censorship Measurement* (right) test sets—Note: X-axes are reversed. The blue lines display the linear regression for each measurement and test list. Correlation coefficients are β=.02 (p=.76) [Satellite, CC], β=-.33/-.04 (p=.0/.04) [Satellite, CR], β=-.76/-.04 (p=.00/.44) [Satellite, CM], β=-.07 (p=.09) [Echo, CC], β=-.38/-.17 (p=.00/.01) [Echo, CR], β=-1.55/-1.25 (p=.00/.01) [Echo, CM]. Negative coefficients reflect a higher average blocking rate when a country is qualitatively rated as more restrictive. The second values, if applicable, show the coefficients after removing influential observations with a Cook's distance above 1. Discard measurements are comparable to the Echo measurements.

Switzerland and Croatia. Both countries are not typically known for online censorship but many DNS probes containing websites from our COVID-containing and COVID-related domains appear to be filtered in networks that are in these countries. Second, we look at Canada as another unexpected country for which we found high average blocking rates for the COVID-containing test list using Quack measurements. Finally, we summarize results for some of the other countries in which we found high censorship: China, Iran, and Kazakhstan.

Switzerland According to "Varieties of Democracies" [11], Switzerland can be considered one of the freest countries when it comes to Internet freedom. However, our DNS measurements in Switzerland detect a high average blocking rate (4.32%) in particular for keywords in our COVID-containing test list.

How is the blocking spread out? We performed measurements to 922 Satellite vantage points spanning 34 ASes in Switzerland. As shown in Figure 3, the number of domains blocked differs by vantage point, even within the same AS.

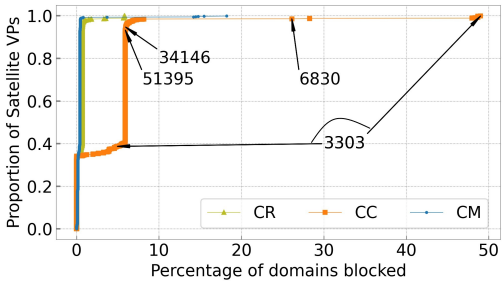

Fig. 3. CDF showing blocking of COVID-related, COVID-containing, and Censorship Measurement domains in Switzerland—The AS numbers of the vantage points experiencing filtering of COVID-containing domains are annotated.

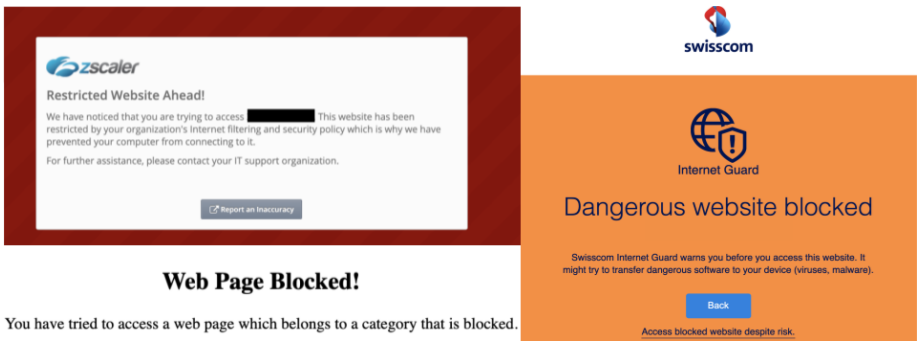

Fig. 4. Switzerland blockpages—The blockpages in the top left and right are ISP blockpages, while the blockpage in the bottom left is a known blockpage of the web filter Fortinet [44].

Six out of 34 ASes have at least one vantage point observing blocking of keywords from our COVID-related and COVID-containing test lists. The AS with the largest amount of vantage points, AS3303 (846 vantage points), observes high blocking of content related to the pandemic. 82 out of the 152 domains in the COVID-containing test list are filtered in our probes to at least one vantage point in this AS. 599 vantage points in AS3303 observe blocking of at least one keyword from the COVID-related list. This AS is the second largest in Switzerland according to Censys [15].

What is the censored response? We find that 601 out of the 607 (99%) vantage points experiencing blocking in Switzerland observed five distinct IP addresses for DNS resolutions of filtered domains, all of which hosted a visible blockpage. Ten vantage points in AS3303 responded with two IP addresses hosting the blockpage shown in Figure 4 (top left) for 74 domains from our COVID-containing list. Interestingly, only domains from the COVID-containing list are resolved to these IPs. Two vantage points in AS3303 and one in AS6830 observed DNS

Table 3. Top 10 filtered domains in Switzerland (DNS).

Domain	Category	% of VPs	Domain	Category	% of VPs
www.covid-19.uk.com	Phishing	66.44	covid-19-stats.info	Phishing	65.74
coronavirus-realtime.com	Malicious	66.40	coronavirus.zone	Malicious	65.41
covid19graph.work	Phishing	66.36	coronavirus-map.com	Phishing	65.36
www.covid19ireland.com	Phishing	66.22	coronastats.net	Malicious	63.60
www.covid19maps.info	Phishing	65.83	coronavirusfrance.org	Phishing	1.58

resolutions to an IP address hosting the blockpage shown in Figure 4 (bottom left), which has previously been identified as one of the blockpages of the web filter Fortinet [44]. The other two IP addresses hosting the blockpage shown in Figure 4 (right) are observed for nine domains from the COVID-containing list, but these are observed in a large number of vantage points (481 & 108).

What are the websites that are blocked? We explore the top websites from our COVID-related test list that are blocked in our probes to DNS resolvers in Switzerland. As shown in Table 3, most of the large-scale blocking in Switzerland seems to be for protecting users from Phishing or Malicious websites (as categorized by Fortiguard's Web Filter service [19]). All of the top 10 websites also fall in our COVID-containing set, primarily contain COVID-19 specific information, and are all categorized as websites containing security threats.

Croatia Croatia has the second-highest average blocking of COVID-containing domains (2.39%) in Satellite measurements. Similar to Switzerland, Croatia is generally considered as free in the online space, and hence such high levels of filtering of COVID-19 specific content deserve scrutiny. We perform measurements to 12 vantage points in Croatia, spread across six ASes. The vantage point that observes the highest rate of blocking for both COVID-containing (28.66%) and COVID-related (3.49%) domains is located in AS5391 and observes redirection to the Fortinet blockpage shown in Figure 4 (bottom left) when tested with 43 domains from the *COVID-containing* test list. One other domain from the COVID-related test list is also blocked (`droneinfini.fr`). Similar to our observation in Switzerland, we observe high blocking of Phishing and Malicious websites in Croatia.

Canada We find significant amounts of application layer keyword filtering in Canada. On average, we see 0.82% and 1.42% blocking of COVID-containing domains in Quack Echo and Quack Discard measurements respectively.

How is the blocking spread out? Quack collected measurements from 201 Echo vantage points distributed across 52 different ASes and 109 Discard vantage points in 34 ASes in Canada. Thirteen Echo vantage points observe blocking of at least one domain from the COVID-related test list, and twelve of these also observe blocking of at least one domain from the COVID-containing test list.

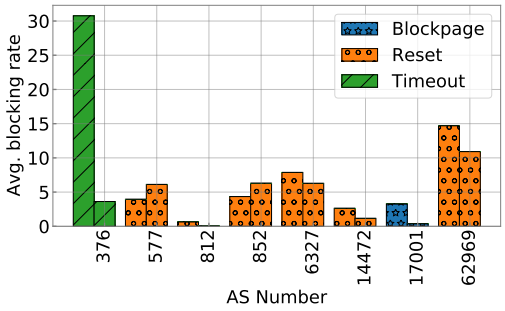

Fig. 5. Blocking distribution in Quack Echo measurements in different ASes in Canada—The left bar shows the average blocking rate of the COVID-containing list and the right bar shows the average blocking rate of the COVID-related list.

Table 4. Top 5 blocked domains in Canada (Application Layer).

Quack Echo			Quack Discard		
Domain	**Category**	**% of VPs**	**Domain**	**Category**	**% of VPs**
covid-19.uk.com	Phishing	3.93	covid19stats.global	Phishing	5.50
covid19stats.global	Phishing	2.86	coronastats.net	Malicious	4.63
covid-19incanada.com	Business	2.84	covid-19canada.com	Business	4.63
covid19uk.live	Reference	2.82	covid-19ireland.com	Not Rated	4.63
www.covid19-maghreb.live	Phishing	2.81	coronavirus-realtime.com	Malicious	4.63

Figure 5 shows the amount of blocking across different ASes in our Echo measurements. AS376 observes the highest amount of COVID-containing blocking. Most of the ASes show considerable blocking of both COVID-containing *and* COVID-related domains. Six Discard vantage points observe blocking of at least one domain from the COVID-related and COVID-containing test lists. In our Discard measurements, we observe similar rates of blocking as in Figure 5 in three ASes: AS376, AS812, and AS17001.

What is the censored response? Figure 5 also shows the type of blocking that is performed in the different ASes. While a majority of ASes in Echo measurements inject reset packets, probes to vantage points in AS376 experience connection timeouts, and the vantage point in AS17001 observes a blockpage. The blockpage explicitly mentions that the content has been blocked because it might contain malicious content. We see similar type of blocking in Discard for the ASes performing blocking. Thus, users in Canada observe different censored responses based on the network they connect to.

What websites are blocked? Similar to Switzerland and Croatia, a significant proportion of the top blocked websites in Canada may be targeted because they are being perceived to be phishing or malicious (See Table 4). All of the top five blocked domains in Canada (in both Echo and Discard measurements) are

COVID-containing domains. Analyzing measurements in the AS that observes the highest amount of COVID-containing blocking, AS376 (RISQ-AS), which consists of 4 vantage points, we see that all of the 48 distinct domains filtered are COVID-containing domains. This AS observes the same blocking pattern in Discard measurements as well. The five distinct domains AS17001 observes to be blocked also belong to our COVID-containing test-list.

Other Countries Countries which typically experience high levels of Internet censorship such as China and Iran also observe high blocking of both *COVID-containing* and *COVID-related* domains (see Table 2). While blocking in these countries may not be strictly related to or caused by the pandemic, it may still hinder users trying to obtain valuable news about the pandemic.

Iran We performed Quack measurements to 39 Echo vantage points spread across 17 ASes and 11 Discard vantage points spread across 9 ASes in Iran. In both our Echo and Discard measurements, we observe high blocking of popular news websites (e.g., `www.huffpost.com`) and social networking websites (e.g.,`www.facebook.com`). In most cases, the blocked response is either a well-known blockpage [44] or a connection timeout. We also performed DNS measurements to 395 vantage points across 61 ASes in Iran, and observe similarly high blocking of popular websites in the COVID-related test list. Some domains in the COVID-containing test list are also blocked (e.g., `coronavirusireland.ie`), indicating filtering policies against websites with pandemic information.

China Our measurements to 1,417 Echo vantage points (in 70 ASes) and 337 Discard vantage points (in 33 ASes) in China observe large-scale blocking of popular news and media websites and Google services. In China, the majority of blocked responses are connection resets. DNS measurements to 4,279 Satellite vantage points in 60 ASes also show similarly high blocking of COVID-related websites containing news. While the blocking of *COVID-containing* domains forms a smaller proportion of the blocking of *COVID-related* domains, some networks block COVID-specific websites such as `covid19japan.com`.

Kazakhstan and Ecuador We also observe significant blocking of both *COVID-containing* (2.23%) and *COVID-related* (0.57%) domains in DNS measurements to Kazakhstan. In this case, domains are resolved to state or Internet Service Provider (ISP) blockpages. In Ecuador, we observe significant blocking of both sets of domains using reset injection in application layer measurements (CC 2.5%, CR 2.29%).

4.3 What categories of COVID-related websites are blocked?

Figure 6 shows the distribution of blocking for different countries in five categories: Business, Government and Legal Organizations, Health and Wellness, News and Media, and Phishing. As observed in the previous section, Phishing

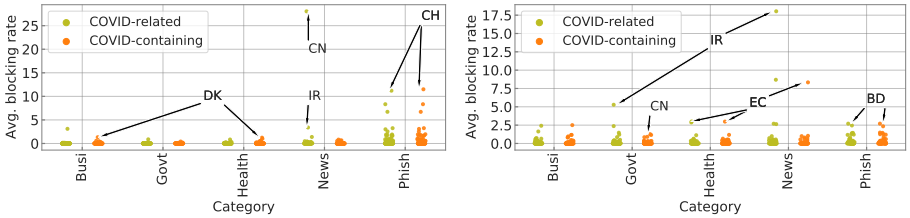

Fig. 6. Blocking distribution across categories—Left: Satellite, right: Quack Echo. The categories were obtained using FortiGuard. Each point on the graph represents a country and the top blocking countries for certain categories have been labeled.

websites observe significant amount of blocking, in both DNS and Application-layer measurements. Since 24.3% of the COVID-containing test list is categorized as Phishing, this blocking appears to be considerably new (since the pandemic started) and COVID-specific. Many news and government websites, which may contain important COVID-19 information, are blocked in our measurements to countries which are known to perform Internet censorship.

4.4 Do COVID-related websites perform phishing?

Throughout our findings, we observe a high blocking rate of many COVID-containing and COVID-related websites that are characterized by the Fortiguard classification service as phishing. Particularly in Switzerland and Croatia, we find the use of the Fortinet web filter (see Figure 4), which also uses the Fortiguard classification service for blocking dangerous websites.

Given the increased prevalence of phishing during the pandemic [30], such blocking is not surprising. However, it is important that websites containing factual information about COVID-19 without any security threats are not blocked by mistake. To better understand whether the 46 websites blocked by the Fortinet web filter in Switzerland and Croatia are actually security threats, we compare it with three freely available URL classifying services: Checkphish [7] (an online website that checks for signs of phishing in an URL), Palo Alto Networks [29] and WatchGuard [50]; two popular web filters [44].

We report our detailed results in Appendix B. We observe substantial differences in the categorization. Fortiguard classifies 42 out of the 46 websites tested as security risks with the tags “Phishing” or “Malicious”. Palo Alto networks only classifies eight of the 46 as “High risk,” “Malware” or “Medium risk.” Watchguard classifies 17 of the 46 domains as “Compromised”, “Suspicious”, “Elevated Exposure” or “Malicious.” Six websites are considered risky by both WatchGuard and Palo Alto Networks. Checkphish did not classify any websites as Phishing.

We also manually determine whether each of these websites contain evident security risks. Three of the authors individually visited each of these websites, categorized them, and reached a consensus. Other than `coronavirus`

`-monitor.ru`, which has an insecure looking popup box where card information can be entered, none of the other websites seem to contain visible security threats. Note that the manual classification does not consider the legitimacy of the data. However, many of the websites list their sources (some common ones are Johns Hopkins University dashboard [23] and government websites) and also warn users that there could be inaccuracies in the data. These findings highlight an important issue with web filter-based censorship—the lack of proper auditing and incorrect categorizations of websites may lead to high amounts of unnecessary blocking for thousands of users, given that these web filters are often used by organizations, ISPs, and governments for blocking dangerous websites [35,44].

5 Discussion

Implications and Future Work Due to the COVID-19 "infodemic", there has been a significant shift in the priorities of many countries throughout the world—the challenge at hand has been achieving a balance between allowing citizens to access important information while also protecting them from harmful misinformation. Our study shows that the large-scale use of URL filtering services may be inadvertently tipping the scale in the wrong direction. While URL classifying and filtering services are known to contain mistakes [33, 47], our results indicate that highly searched (and potentially harmless) domains are being blocked in several countries due to these errors. There is a serious lack of transparency surrounding the decisions made by filters and large discrepancies from filter to filter, a concern also echoed by recent work [47]. These issues make detailed auditing of such services necessary. We advocate for further research into the mechanics of filters and their categorization techniques, and for third-parties to monitor and track filtering policies that affect a large number of users, such as ISP and country-level deployments [44].

In cases where COVID-19 related censorship is more intentional, countries could be using genuine reasons supporting the need for information control during the pandemic as a facade for restricting information and continuing censorship in the long run for unrelated reasons [36, 52]. Future work should track the overflow of censorship policies enacted during the pandemic over time to prevent unnecessary loss of access. Moreover, more in-depth analyses of the contents of a website will help to determine the reason behind filtering. A recent report [55] highlights that content critical of China's handling of COVID-19 have been reported to be taken down in China, for instance.

Finally, while this study focuses on the filtering of websites related to popular searches of factual COVID-19 information, there is a possibility that misinformation related to COVID-19 is blocked more restrictively. Future work can use our measurement tools to monitor websites that are more likely to contain misinformation. In addition, we do not account for website filtering performed by search engines themselves; whereas top factual results are rarely suppressed by search engines, future work studying misinformation needs to consider whether search engine censorship is a significant contributor to information unavailabil-

ity. Search results obtained from other search engines such as Yandex or Baidu could also be incorporated in test-lists to allow for more comprehensive findings.

Limitations When assembling the input lists, we used results from search crawls conducted in nine geo-distributed countries. Our censorship measurements included several countries that were not included in this list and thus, it is possible that there are resources local to these countries that are being filtered. Moreover, by only using the top ten results from each search, we potentially miss measuring the filtering of less popular websites, which we leave for future work.

Even though we run measurements from a large number of vantage points around the globe, our vantage points do not have the granularity required to detect all blocking. Moreover, only a handful of vantage points are available in some countries, and hence our observations may be limited to a specific network or region. Quack sends measurements to port 7 and port 9 and therefore may miss censorship that is only applied to traffic on port 80 or port 443. In addition, the Quack Discard technique cannot detect censorship that only affects outbound traffic. However, studies have shown that such censorship is difficult to perform, so it is unlikely to substantially alter our test results [48]. There is also the possibility that some censors apply mechanisms to evade our detection, although we are not aware of any such measures to-date. Finally, the Maxmind geolocation database we used is known to have inaccuracies [20].

6 Conclusion

In this paper, we have explored the global extent of censorship related to the pandemic using a test list of popular COVID-19 websites and remote censorship measurement techniques. We find generally low levels of blocking related to the pandemic. However, we observe that commercial URL filtering services deployed in countries such as Switzerland and Canada mistakenly consider many COVID-19-related websites as containing phishing threats and block them. When censors engage in blocking of this kind, be it well-intentioned or purely suppressive, it has the potential to cut off this vital flow of information. As an online community, we must advocate for stricter auditing of filtering practices for ensuring that essential information is available to every person that needs it.

7 Acknowledgments

The authors thank the shepherd Philipp Winter and the reviewers for their constructive feedback. We also thank Prerana Shenoy for her help with data analysis. This work was supported in part by research credits from Google.

References

1. S. Afroz and D. Fifield. Timeline of Tor censorship, 2007. http://www1.icsi.berkeley.edu/~sadia/tor_timeline.pdf.

2. Alexa Internet, Inc. Alexa Top 1,000,000 Sites. http://s3.amazonaws.com/alexa-static/top-1m.csv.zip.
3. H. Allcott, M. Gentzkow, and C. Yu. Trends in the diffusion of misinformation on social media. *Research & Politics*, 2019.
4. Anonymous. Towards a comprehensive picture of the Great Firewall's DNS censorship. In *USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2014.
5. S. Aryan, H. Aryan, and J. A. Halderman. Internet censorship in Iran: A first look. In *USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2013.
6. J. S. Brennen, F. Simon, P. N. Howard, and R. K. Nielsen. Types, sources, and claims of covid-19 misinformation. *Reuters Institute*, 2020.
7. CheckPhish. Url Scanner to Detect Phishing in Real-time — CheckPhish. https://checkphish.ai/.
8. C. Cheng, J. Barceló, A. S. Hartnett, R. Kubinec, and L. Messerschmidt. Covid-19 government response event dataset (coronanet v. 1.0). *Nature Human Behaviour*, 2020.
9. M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala. The covid-19 social media infodemic. *arXiv preprint arXiv:2003.05004*, 2020.
10. Citizen Lab. Block test list. https://github.com/citizenlab/test-lists.
11. M. Coppedge, J. Gerring, C. H. Knutsen, S. I. Lindberg, J. Teorell, D. Altman, M. Bernhard, M. S. Fish, A. Glynn, A. Hicken, et al. V-dem codebook v10, 2020. https://www.v-dem.net/media/filer_public/28/14/28140582-43d6-4940-948f-a2df84a31893/v-dem_codebook_v10.pdf.
12. M. Crete-Nishihata, J. Dalek, J. Knockel, N. Lawford, C. Wesley, and M. Zhou. Censored contagion ii: A timeline of information control on chinese social media during covid-19. https://citizenlab.ca/2020/08/censored-contagion-ii-a-timeline-of-information-control-on-chinese-social-media-during-covid-19/, 2020.
13. A. L. Dahir. Internet shutdowns are costing African governments more than we thought. https://qz.com/1089749/internet-shutdowns-are-increasingly-taking-a-toll-on-africas-economies/.
14. A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé. Analysis of country-wide internet outages caused by censorship. In *ACM Internet Measurement Conference (IMC)*, 2011.
15. Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman. A search engine backed by Internet-wide scanning. In *Proceedings of the 2015 ACM SIGSAC Conference on Computer and Communications Security*, 2015.
16. Z. Durumeric, E. Wustrow, and J. A. Halderman. ZMap: Fast internet-wide scanning and its security applications. In *USENIX Security Symposium*, 2013.
17. R. Ensafi, P. Winter, A. Mueen, and J. R. Crandall. Analyzing the Great Firewall of China over space and time. *Proceedings on Privacy Enhancing Technologies (PETS)*, 2015.
18. E. Ferrara. What types of COVID-19 conspiracies are populated by twitter bots? *First Monday*, 2020.
19. FortiNet. Fortiguard labs web filter. https://fortiguard.com/webfilter.
20. M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos. A look at infrastructure geolocation in public and commercial databases. In *ACM Internet Measurement Conference (IMC)*, 2017.
21. Coronavirus search trends - google trends. https://trends.google.com/trends/story/US_cu_4Rjdh3ABAABMHM_en.

22. J. Jerreat. Coronavirus the new scapegoat for media censorship, rights groups say. https://www.voanews.com/press-freedom/coronavirus-new-scapegoat-media-censorship-rights-groups-say, 2020.
23. Johns Hopkins University, Coronavirus Resource Center. Covid-19 dashboard by the center for systems science and engineering (csse) at johns hopkins university (jhu). https://coronavirus.jhu.edu/map.html.
24. Kyaw, Phyu Phyu and Xynou, Maria and Filastò, Arturo. Myanmar blocks "fake news" websites amid covid-19 pandemic. https://ooni.org/post/2020-myanmar-blocks-websites-amid-covid19/.
25. J. Lachapelle, A. Lührmann, and S. F. Maerz. An update on pandemic backsliding: Democracy four months after the beginning of the covid-19 pandemic. *V-Dem Institute: Policy Brief*, 2020.
26. R. MacKinnon. China's censorship 2.0: How companies censor bloggers. *First Monday*, 2009.
27. MaxMind. https://www.maxmind.com/.
28. Ofcom. Half of uk adults exposed to false claims about coronavirus. https://www.ofcom.org.uk/about-ofcom/latest/media/media-releases/2020/half-of-uk-adults-exposed-to-false-claims-about-coronavirus, 2020.
29. Palo Alto Networks. Test a site. https://urlfiltering.paloaltonetworks.com/.
30. PC Magazine. Phishing attacks increase 350 percent amid covid-19 quarantine. https://in.pcmag.com/privacy/135635/phishing-attacks-increase-350-percent-amid-covid-19-quarantine.
31. P. Pearce, R. Ensafi, F. Li, N. Feamster, and V. Paxson. Augur: Internet-wide detection of connectivity disruptions. In *IEEE Symposium on Security and Privacy (S&P)*, May 2017.
32. P. Pearce, B. Jones, F. Li, R. Ensafi, N. Feamster, N. Weaver, and V. Paxson. Global measurement of DNS manipulation. In *USENIX Security Symposium*, 2017.
33. P. Peng, L. Yang, L. Song, and G. Wang. Opening the blackbox of virustotal: Analyzing online phishing scan engines. In *ACM Internet Measurement Conference (IMC)*, 2019.
34. R. Radu. Fighting the 'infodemic': Legal responses to COVID-19 disinformation. *Social Media + Society*, 2020.
35. R. Ramesh, R. Sundara Raman, M. Bernhard, V. Ongkowijaya, L. Evdokimov, A. Edmundson, S. Sprecher, M. Ikram, and R. Ensafi. Decentralized Control: A Case Study of Russia. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2020.
36. Reporters Without Borders. Middle east governments clamp down on coronavirus coverage, 2020. https://rsf.org/en/news/middle-east-governments-clamp-down-coronavirus-coverage.
37. University of Oregon Route Views Project. www.routeviews.org.
38. L. Ruan, J. Knockel, and M. Crete-Nishihata. Censored contagion: How information on the coronavirus is managed on chinese social media. https://citizenlab.ca/2020/03/censored-contagion-how-information-on-the-coronavirus-is-managed-on-chinese-social-media/, 2020.
39. W. Scott, T. Anderson, T. Kohno, and A. Krishnamurthy. Satellite: Joint analysis of CDNs and network-level interference. In *USENIX Annual Technical Conference (ATC)*, 2016.
40. SeleniumHQ Browser Automation. www.selenium.dev.
41. L. Singh, S. Bansal, L. Bode, C. Budak, G. Chi, K. Kawintiranon, C. Padden, R. Vanarsdall, E. Vraga, and Y. Wang. A first look at COVID-19 information and misinformation sharing on twitter, 2020.

42. R. Sundara Raman, L. Evdokimov, E. Wustrow, A. Halderman, and R. Ensafi. Investigating Large Scale HTTPS Interception in Kazakhstan. In *Internet Measurement Conference (IMC)*. ACM, 2020.
43. R. Sundara Raman, P. Shenoy, K. Kohls, and R. Ensafi. Censored Planet: An Internet-wide, Longitudinal Censorship Observatory. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2020.
44. R. Sundara Raman, A. Stoll, J. Dalek, A. Sarabi, R. Ramesh, W. Scott, and R. Ensafi. Measuring the Deployment of Network Censorship Filters at Global Scale. In *Network and Distributed System Security Symposium (NDSS)*, 2020.
45. The Tor Project. OONI: Open observatory of network interference. https://ooni.torproject.org/.
46. University of Toronto. Citizen Lab. https://citizenlab.ca/.
47. P. Vallina, V. Le Pochat, Á. Feal, M. Paraschiv, J. Gamba, T. Burke, O. Hohlfeld, J. Tapiador, and N. Vallina-Rodriguez. Mis-shapes, mistakes, misfits: An analysis of domain classification services. In *ACM Internet Measurement Conference (IMC)*, 2020.
48. B. VanderSloot, A. McDonald, W. Scott, J. A. Halderman, and R. Ensafi. Quack: Scalable remote measurement of application-layer censorship. In *USENIX Security Symposium*, 2018.
49. A. Vyas, R. Sundara Raman, N. Ceccio, P. M. Lutscher, and R. Ensafi. Investigating Filtering of COVID-19 Websites, 2020. https://censoredplanet.org/covid.
50. WatchGuard. See a site's content category. https://www.watchguard.com/help/docs/help-center/en-US/Content/en-US/Fireware/services/webblocker/site_categories_see_websense_c.html.
51. P. Winter and S. Lindskog. How the Great Firewall of China is blocking Tor. In *USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2012.
52. J. Wiseman. European media freedom suffers under COVID-19 response, 2020. https://ipi.media/european-media-freedom-suffers-covid-19-response/.
53. World Health Organization. Coronavirus disease (COVID-19) pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019, 2020.
54. X. Xu, Z. M. Mao, and J. A. Halderman. Internet Censorship in China: Where Does the Filtering Occur? In *International Conference on Passive and Active Network Measurement (PAM)*, 2011.
55. R. Zhong, P. Mozur, J. Kao, and A. Krolik. No 'Negative' News: How China Censored the Coronavirus, 2020. https://www.nytimes.com/2020/12/19/technology/china-coronavirus-censorship.html.
56. J. Zittrain and B. Edelman. Internet filtering in China. *IEEE Internet Computing*, 2003.

A Search Engine Crawl Keywords

Table 5 shows the list of prefix and suffix combinations used to construct the keywords used for our search engine crawls.

B Classifier and Manual Categorizations

Table 6 shows the results of the categorization of the 46 domains blocked by the Fortinet filter in Switzerland and Croatia using different categorization tools.

Table 5. Keyword permutations used for search engine crawls. Three terms (corona virus, covid virus, and covid-19 virus) are excluded from the table.

Prefix	Suffix
coronavirus, covid, covid-19	usa, ireland, uk, britain, india, canada, singapore, korea, japan, australia, germany, france, update, news, worldometer, deaths, victims, map, live, infections, stats, toll, death toll, vaccine, dead, <empty>

Table 6. URL Classifier and Manual Categories—Note that each tool uses different category labels. Checkphish(CP) only categorizes websites as yes/no for phishing. Palo Alto Networks (PAN) classifies websites according to their content and risk level but only the risk category is shown here. Our manual classification indicates yes/no for whether the websites contained evident phishing threats. NTE stands for Navigation Time Exceeded, E stands for error and NC stands for not categorized.

Website	**FortiGuard**	**CP**	**PAN**	**WatchGuard**	**Manual**
canadacovid.ca	Phishing	no	Low	NC	no
co19stats.com	NC	no	Low	NC	no
coronastats.net	Malicious	no	Malware	Malicious Web Sites	no
coronavictimes.net	Phishing	no	Low	Elevated Exposure	no
coronavirus-global.com	Phishing	no	Low	Sports	no
coronavirus-map.com	Phishing	NTE	High	Malicious	no
coronavirus-map.org	Phishing	no	Low	Health	no
coronavirus-monitor.com	Phishing	no	Low	Business and Economy	no
coronavirus-monitor.ru	Malicious	NTE	Medium	Health	yes
coronavirus-realtime.com	Malicious	NTE	Malware	Compromised	no
coronavirus.zone	Malicious	no	Malware	Malicious	no
coronavirusfrance.org	Phishing	no	Low	Elevated Exposure	no
coronavirusireland.ie	Phishing	no	Low	News and Media	no
coronavirusmap.co.uk	Phishing	no	Low	NC	no
coronavirusstatistics.org	Phishing	NTE	Low	NC	no
coronavirusupdate.me	Phishing	no	Low	Shopping	no
coronavirususamap.com	Phishing	no	Low	NC	no
covid-19-fr.fr	Phishing	no	Low	Health	no
covid-19-stats.info	Phishing	no	Malware	Malicious	no
covid-19.uk.com	Phishing	no	Low	NC	no
covid-19ireland.com	NC	E	Low	Elevated Exposure	no
covid-japan.com	Phishing	no	Low	News and Media	no
covid-live.net	Phishing	no	Low	Elevated Exposure	no
covid-stats.net	NC	no	Low	Elevated Exposure	no
covid19-uk.co.uk	Phishing	no	Low	Elevated Exposure	no
covid19dashboard.live	Phishing	no	Low	NC	no
covid19graph.work	Phishing	no	Low	Malicious Web Sites	no
covid19live.org	Phishing	no	Low	Elevated Exposure	no
covid19statistics.org	Phishing	no	Low	Government	no
covid19stats.global	Phishing	no	High	Malicious	no
covid19video.com	Phishing	E	Low	NC	no
droneinfini.fr	Phishing	no	Low	NC	no
koronavirus-today.ru	Phishing	no	Low	NC	no
map-covid-19.com	Phishing	NTE	Low	Reference Materials	no
ru.coronavirus-global.com	Phishing	no	Low	Sports	no
wa-daily-covid-19.com	Phishing	no	Low	Elevated Exposure	no
worldcoronavirus.org	Phishing	no	Low	Elevated Exposure	no
worldometers.cc	NC	E	Low	Suspicious content	no
www.coronalive.info	Phishing	NTE	Low	NC	no
www.coronavictimes.fr	Phishing	no	Low	Business and Economy	no
www.coronavirus-india.net	Phishing	no	Low	NC	no
www.covid19-maghreb.live	Phishing	no	Low	NC	no
www.covid19ireland.com	Phishing	no	Malware	Government	no
www.covid19maps.info	Phishing	NTE	Low	NC	no
www.covidstats.com	Phishing	no	Low	NC	no
www.vaccin-coronavirus.fr	Phishing	no	Low	Health	no