Profiting from the Pandemic: Online Supplement

Yung Au, Philip N. Howard, Project Ainita

1. Joan Donovan's framework for actors down the Technology Stack that have roles in moderating content online

Terms of The Technology Stack	Definition
The Open Web (Level 1)	The public web which is accessible by anyone Example of moderation: Individual websites banning users (e.g. banned users on messaging boards)
Platforms (Level 2)	Digital platforms are technology-enabled content intermediaries Example of moderation: The removal and curation of content on platforms (e.g. Facebook removing inappropriate posts)
Cloud Services (Level 3)	Services that provide access to applications/resources on demand via remote cloud computing servers so clients do not need to own such infrastructures themselves Example of moderation: Providers refusing services to clients who hosts illegal or stolen content (e.g. Google refusing services to 8chan)
Content Delivery Networks (CDNs) (Level 4)	Services that offer geographically distributed network of servers to provide rapid delivery of Internet content Example of moderation: Providers refusing services to clients who hosts illegal or stolen content (e.g. Cloudflare refusing services to 8chan)
Registrars (Level 5)	An accredited provider that handles website name registrations Example of moderation: Content decisions have been rarer on registrar levels but registrars have denied services for things such as trademark infringement or blacklisted clients by government order.
Internet Service Providers (ISP) (Level 6)	Any company that provides some sort of connectivity/access to the Internet, such as your mobile phone operator or home ISP Example of moderation: Moderation here includes blacklisting websites for piracy and selectively throttling bandwidth to certain content (e.g. the net neutrality fight).

Source: Summarized from Joan Donovan, 'Navigating the Tech Stack: When, Where and How Should We Moderate Content?', *Centre for International Governance Innovation*, 2019.

2. Sampling

Sampling: This memo uses a small-to-medium sample of websites to examine and triangulate several open source tools that look into the back-ends of domains. The intention with these curated samples is to minimize the possibility of false positives in the sample rather than random samples. Parked, inactive websites, or websites that have returned 404 codes have been removed. 301 redirects have been retained. A mixture of manual and automatic checks was used to determine status codes and relevance of websites. https://httpstatus.io was also used to triangulate status codes.

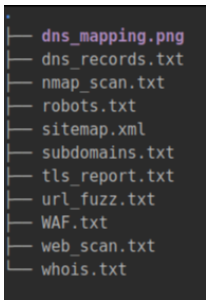
Vetting: Large tech companies such as Google and Facebook are often not entirely transparent about content they takedown or they blacklist (instead, opting for more general comments such as "we have taken down x number of posts and pages"), except for exceptional cases. Part of our methodology thus relied on doing content scans of the sample of websites we have that were to some extent, curated beforehand by another group (either ourselves, Krebs Security, or Artists Against 419) and we matched to what big tech companies had announced they were cracking down on. These organisations that curated website lists also flagged these problematic content to various authorities. For example, Artists Against 419 reports all its scam sites to web hosts and other relevant authorities.

3. Analysis

WHOIS Lookup: A query and response protocol that is widely used for querying databases that store the registered users or assignees of an Internet resource, such as a website name or an IP address block
IP Address: A numerical label assigned to each device connected to a computer network that uses the Internet Protocol for communication
Edge IP Address: Is the first public facing IP address you see. So if using a CDN, this is the first public facing server that is being reached. E.g. If you reach a website from Germany, you might see a Cloudflare IP address from Frankfurt. From there onwards, you will go through Cloudflare's network of servers (which is not public) until you reach the website's origin IP.
Web Hosts: A service that stores all the pages of your website and makes them available to devices connected to the Internet. There are currently two main types of hosting services used by webmasters: web and cloud hosting.
Web based third-party technology: An umbrella term for the wide variety of service and add-ons to websites that provide additional functionalities beyond standard internet technologies such as TCP/IP or HTTP
Trackers: Pieces of software that are meant to collect information of users and user activities. Tracking can be done through browser Cookies, browser fingerprinting, IP tracking and so on.

This memo uses a mixture of Lookup tools including HTML code scans and shell scripting to run batch analysis, as well as triangulating with bulk URL tools such as https://www.bulkseotools.com. North American servers were used to assess IP geolocation.

Scripting tools include RACCOON (https://github.com/evyatarmeged/Raccoon) for the following records:



IP Location Note: It should be noted that the geolocation IP addresses are not the origin IP addresses of websites. Origin IP addresses are usually obfuscated by a variety of means. Furthermore, IP geolocation only provides a partial view of website infrastructure as many websites utilise CDNs which help to conceal the web server's origin identity through processes such as reverse proxy services and load balancers. However, the edge IP address is useful for examining the servers they come into contact with and the cloud services that websites use. Furthermore, IP addresses and routing depends on where your servers are when you are visiting a website.

Third party technology includes: plugins that help website functionality (e.g. mobile responsivity), e-commerce tools (e.g. shopping cart supports and payment apps), and interactive elements (e.g. chat or petition plugins).

Web plugins were triangulated using Ghostery, Wappanalyzer, and Whatruns browser add-ons.

For the TrackerTracker tool: this memo uses a crawl depth of maximum 10 subpages as many trackers are embedded beyond the front page.

On average, websites protesting public health precautions and websites profiteering from the crisis have roughly the same proportions of trackers. About a third of their trackers are for advertising, a third are for analytics, and a third are a mixture of essential trackers and widgets. In contrast, almost two thirds of the trackers on COVID-19 disinformation websites are advertising trackers, revealing how much they depend on advertising as a revenue source. (See table 2)

TABLE 2: DISTRIBUTION OF TRACKERS FROM EACH SAMPLE OF WEBSITES

	Tracker Type	*% of Tracker*
(1) Lockdown Protest	Advertising	32
	Analytics	32
	Essential	9
	Widgets	27
(2) COVID-19 Scam	Advertising	29
	Analytics	29
	Essential	13
	Widgets	29
(3) COVID-19 Disinformation	Advertising	61
	Analytics	20
	Essential	16
	Widgets	3

4. Sample of websites

**please exercise caution when visiting websites as they may contain malicious software (especially flagged scam websites)*

Sample of Protest Websites (n=40):

Domain	**Date Created**
liberatecali.com	5/10/2020
reopencc.com	4/10/2020
reopenarizona.com	4/9/2020
reopeniowa.com	4/8/2020
reopenmn.com	4/8/2020
reopenohio.com	4/8/2020
reopenpa.com	4/8/2020
nevadadeservesbetter.com	5/24/2020
liberatedetroit.com	5/23/2020
resisttheshutdown.com	5/15/2020
americanpatriotrally.com	4/28/2020
liberateamericanow.com	4/24/2020
saveourcountry2020.com	4/20/2020
reopenamericanbusines.com	4/17/2020
reopenalabama.com	4/16/2020
reopensociety.com	4/16/2020
reopentexasnow.com	4/16/2020
reopenmaryland.com	4/15/2020
reopennevadanow.com	4/15/2020
reopennystate.com	4/15/2020

reopenoureconomy.com	4/15/2020
reopensc.com	4/15/2020
reopentx.com	4/15/2020
reopenwi.com	4/15/2020
openthestates.com	4/14/2020
reopenkentucky.com	4/14/2020
reopenkentuckynow.com	4/14/2020
reopennc.com	4/14/2020
reopennh.com	4/14/2020
reopenohionow.com	4/14/2020
americanrevolution2.org	4/13/2020
reopenamerica.info	10/4/2020
reopenamerica.org	3/21/2020
letamericaopen.net	5/2/2020
mymilitia.com	1/26/2016
coronavirustruths.godaddysites.com	11/18/2013
conventionofstates.com	8/24/2012
wehaverights.com	12/1/2011
hawaiideservesbetter.com	2/21/2009
open-nc.us	5/1/2000

Sample of COVID-19 scam websites

Domain	**Date Created**
chinthaiimportandexportco.com	3/12/2020
annaasiasurgicalmasks.com	4/11/2020
medicalelixirkft.com	2/11/2020
healthcarepropharmacy.com	3/10/2020
moencoltd.com	7/18/2020
monthianenterprisecoltd.com	7/15/2020
environsafetyequipments.com	3/7/2020
jbnitrilegloves.com	6/15/2020
kamiencasp-zoo.com	5/6/2020
theblissfulhealth.com	4/6/2020
loverthygiene.com	2/6/2020
esibooterglobal.com	5/23/2020
weisengloves.com	5/23/2020
deptholdingbv.com	5/19/2020
jbmedsuply.com	5/17/2020
abgroupspolkaz.com	5/13/2020
j-manussuntornmedica.com	4/30/2020
meiclinical.com	4/22/2020
thelucidhealth.com	3/28/2020
uspharmaciaspzoo.com	3/22/2020
globalpharmaceuticalssarl.com	3/14/2020

breathfreemask.com	2/24/2020
otisimedicalsupplies.com	2/24/2020
globalfacemask.com	2/21/2020
universalpharmasupplies.com	2/18/2020
surgical-facemaskshop.com	5/1/2020
theluxuryhealing.com	5/1/2020
n95-medical-masks.com	4/1/2020
hallucinogetic.com	11/17/2019
kanchanaperdsann.com	10/30/2019
cbdhempmedic.com	9/16/2019
amscinternationalltd.com	12/9/2019
emgeneral.com	4/9/2019
tropicalprotectivewear.com	8/16/2019
solivartradingafs.com	8/6/2019
globalshippingandlogisticscompany.com	7/5/2019
alabtrades.com	11/4/2019
biskisloriuniplda.com	3/29/2019
thaiglobalshippingcompany.com	8/21/2016
sofian-shipping.com	10/29/2015

Sample of COVID-19 disinformation domains

Domain	**Date Created**
americasfrontlinedoctorsummit.com	7/29/2020
hcqtrial.com	7/28/2020
hcqtrial.com	7/28/2020
covid19refusers.com	5/24/2020
thehealthyamerican.org	5/18/2020
c19study.com	6/5/2020
fakepandemic.com	3/23/2020
c19hcq.com	7/3/2020
plandemicmovie.com	5/1/2020
tierneyrealnewsnetwork.com	6/23/2019
pandemic.news	11/18/2018
nytwatch.com	6/24/2018
nytwatch.com	6/23/2018
wapoop.news	6/23/2018
medicalextremism.com	9/12/2017
infections.news	12/3/2017
channel22news.com	1/15/2017
bigleaguepolitics.com	4/30/2016
newsfakes.com	6/11/2015
vaccineinjurynews.com	6/10/2015
redstatewatcher.com	8/15/2015
scienceclowns.com	4/26/2015

theamericanmirror.com	12/14/2014
healthnutnews.com	11/2/2013
themindunleashed.com	1/18/2013
healthimpactnews.com	1/11/2011
wakingtimes.com	10/28/2011
humansarefree.com	12/22/2010
naturalhealth365.com	11/1/2010
vactruth.com	5/11/2009
zerohedge.com	1/11/2009
greenmedinfo.com	3/18/2009
naturalnewsradio.com	2/12/2008
beforeitsnews.com	9/15/2007
naturalnews.com	2/19/2005
americanthinker.com	4/24/2003
jimbakkershow.com	4/18/2003
westonaprice.org	7/26/1999
wnd.com	9/23/1998
rushlimbaugh.com	7/6/1995