Registrations Related to COVID-19: 18 Months of Data

ICANN Office of the Chief Technology Officer

Siôn Lloyd

OCTO-028
9 November 2021

TABLE OF CONTENTS

This document supports ICANN's strategic goal to improve the shared responsibility for upholding the security and stability of the Domain Name System (DNS) by strengthening DNS coordination in partnership with relevant stakeholders. It is part of ICANN's strategic objective to strengthen the security of the DNS and the DNS root server system (RSS).

This document is part of ICANN's Office of the Chief Technical Officer (OCTO) document series. Please see the OCTO publication page for a list of documents in the series. If you have questions or suggestions on any of these documents, please send them to octo@icann.org.

Executive Summary

After 18 months of inspecting COVID-19-related domain name registrations, for reports of phishing or malware delivery, the Office of the Chief Technology Officer (OCTO) of the ICANN organization (ICANN org), finds that, depending on the strength of evidence required, between 1.8% to 6.1% could be said to be malicious. Importantly, this means that the vast majority of registrations have no strong evidence of misuse, particularly under these two categories of malicious use (activity like fraud or scams are out of the scope of the Domain Name System Security Threat Information Collection and Reporting (DNSTICR) project).

We also see how the volumes of registrations spike at the beginning of a widely-reported event (in this case, the COVID-19 pandemic) and fall to near-background levels after three months. Over the time period analyzed, the various terms we are looking for come in and out of popularity, reflecting how society changed and evolved how it referred to the pandemic.

Finally, we note how many phishing campaigns were observed that tapped into offers of financial relief or official government schemes. Other pages show the sorts of fake login pages that have been commonly seen historically, sometimes with an extra COVID-19-related message.

1 Introduction

A year and a half ago, the world was just coming to understand that COVID-19 was a pandemic of global proportions. Sadly, bad actors who use newsworthy events to drive traffic to their malicious websites were also waking up to it, and we started seeing many reports in the popular press and elsewhere of phishing attacks, in particular using “COVID” and surrounding terms in their domain names. Many of these reports had large headline-grabbing numbers yet little or no information about collection methodology, definitions, and so on, meaning that they could not be easily verified.

In response to these events, several groups began making available "threat intelligence," that is, information and data on reported or observed security threats related to these domains. Some groups, like the COVID-19 Cyber Threat Intelligence League and the COVID-19 Cyber Threat Coalition, also began to take action to counter these malicious actors.

ICANN org began contributing to this COVID-19 anti-abuse effort, using our knowledge and expertise to put actionable intelligence into the hands of those able to disrupt malicious campaigns. To this end OCTO developed the Domain Name Security Threat Information Collection and Reporting (DNSTICR) project to report COVID-19-related Domain Name System (DNS) security threats. We began filtering lists created from zone files and enriching them with data from external sources to separate domain names that may be malicious from the majority that are not.

During this project we have been careful not to add more noise to those in a position to take action because they were already swamped with information. While we aim to provide data that allows a rapid assessment of a domain's status, we also need a high level of confidence in the assessment. Without both, we may end up doing more harm than good.

OCTO is producing reports on recent domain registrations that we believe are using the COVID-19 pandemic for phishing or malware campaigns. These reports, which are shared with the responsible parties (the registrars), contain the evidence that leads ICANN org to believe the domains are being used maliciously, along with other background information to help the responsible parties determine the correct course of action.

2 Process Detail

To generate these reports, we examine available generic top-level domain (gTLD) zone files. These files allow us to see newly delegated registrations (although our report generation process can technically receive input from any source of domains, hostnames or full URLs).

These files are filtered for new entries that contain words like "COVID," "corona," "pandemic," and other terms related to the pandemic. We have modified this list of terms through the course of the pandemic in order to include new ones, such as pharmaceutical companies involved in vaccine production. We also include translations and non-Latin character versions to maximize our global coverage, although it should be noted that we are not analyzing country-code top-level domains (ccTLDs), and so will be missing that part of the picture. When new terms are added, a historic search is made to locate any registrations matching those terms from the start of the project.

The resulting set of domain names forms our list of those we have identified as potentially related to the COVID-19 pandemic. This list though is still just data and is not actionable as it will include both benign and malicious domains. We need to further refine the list to improve our confidence in the contents.

The next step is to take our filtered list and look across a number of third-party threat intelligence sources for indications of the domains being used for phishing or malware distribution. We want any evidence that we see to be verifiable by the parties receiving our reports, therefore, the sources must have public access to some level of the intelligence. We start by using four sources, Virus Total, AlienVault OTX, Phishtank, and Google Safe Browsing. Our report generation process is designed to be extensible so that threat intelligence sources can be added or removed.

The data provided by these sources can suggest a domain is malicious, although in most cases we find little to no evidence of the types of activity that we are looking into. Lack of evidence can be ascribed to:

- The domains being benign
- The domains being "parked," that is speculatively registered with the intent to later sell at a profit or generate revenue via advertising
- Young domains that have not yet been used maliciously or the behavior has not yet been observed
- Domains being used for malicious purposes other than phishing or malware distribution, like fraud or scams
- We are starting from lists of domain names and not full URLs. Unless the malicious behavior is evident from the homepage we may not see it

For some of the sources above it may be the case that, if we were to look again after a period of time, new evidence will be available and that the reputation of the domain will change. In order to allow for this, we periodically retest domains.

As said above, we are specifically looking for malware and phishing, and have chosen sources that focus on these areas. As a result, the malicious activity that we see is largely within these two categories. However, it is not uncommon for domains to exist in multiple categories, because different collection methods employed by threat intelligence providers detect different aspects of a malicious campaign. Also, categorization is not an exact science and can be open to interpretation. Therefore, when we use the term "malicious" here we are referring to phishing or malware. Of course, this categorization does not preclude a domain from appearing on a spam or other undesirable behavior list if it is also used in those ways. However, we will only see the domains when they are reported for the two specific types of activity that we are interested in as part of the project.

When credible evidence of malicious activity is found, we proceed to gather more information about the domains in line with the reporting requirements specified by registrars in the Guide to Registrar Abuse Reporting. This reporting information includes the registrar (and abuse contact in particular), hosting information, and a screenshot if appropriate. This information is gathered to help those receiving our reports make the decision on whether or not to take action (such as suspension) against the domain.

When we say credible evidence we mean that several independent, trusted reports of malicious activity can be found. There are several reasons why some pieces of evidence may not be considered suitable for our particular use case, for example some data feeds report domains with a high probability of being malicious, these are often associated with the use of machine-learning techniques. These probabilistic scores are generated from characteristics of the domain in ways which can be hard to determine, even if we had access to the data and models used, and it is hard for us to determine the accuracy of such a data feed. This does not mean that the data is not valuable or relevant in a different context, but we are looking for direct observations that can be reported. We also have to be careful as some open-source data finds its way into multiple feeds with no clear indication of whether it has been independently verified each time. This means that what looks like multiple observations may in fact be a single observation being reported multiple times.

If we see no evidence of maliciousness, or the evidence that we see is insufficient, we place the domain in a queue and re-examine it seven days later. We do this for a maximum of four weeks after the first observation for each domain. This delay and extended period of observation means that we often see domains with sufficient evidence of maliciousness, however, they have already been dealt with by the registrar or hosting entity. Often the domains no longer resolve or if they do, they resolve to a “suspended” page, a “registrant verification pending” page or a “suspected phishing ahead” page. We do not send reports for these domains as they are no longer able to do harm.

The flowchart shown in Figure 1 summarizes the report generation process.

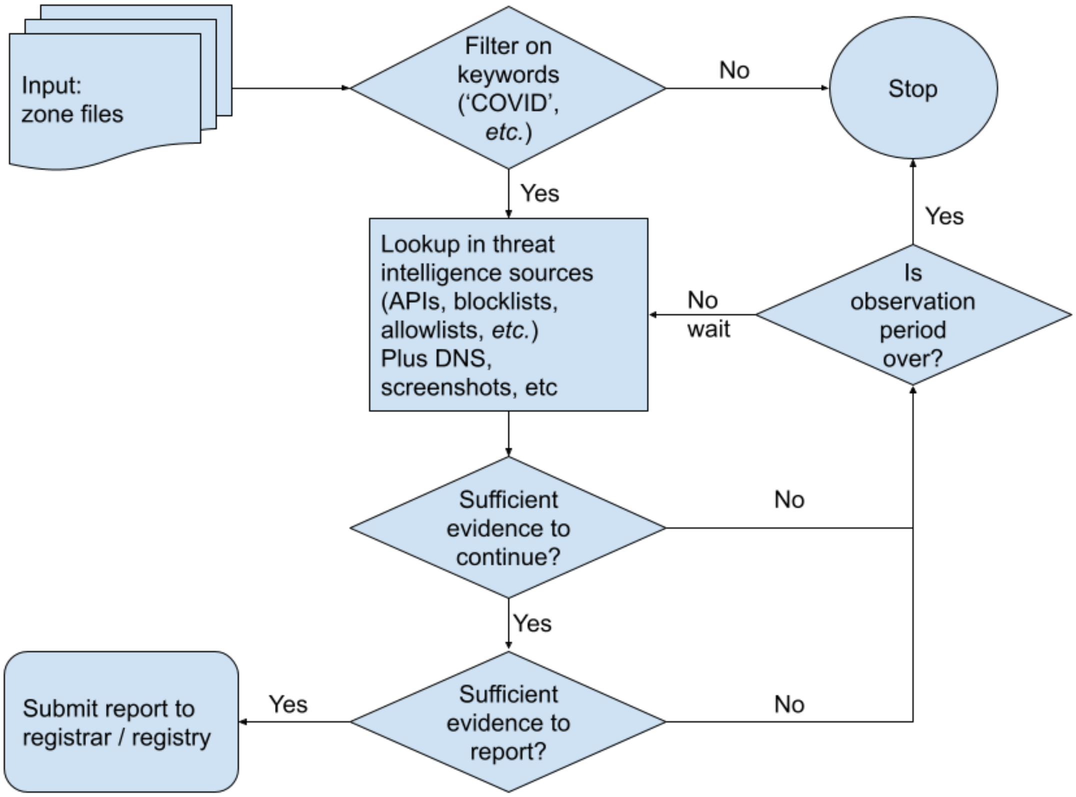

Input:
zone files
Filter on keywords ('COVID', *etc.*)
No
Stop
Yes
Yes
Lookup in threat intelligence sources (APIs, blocklists, allowlists, *etc.*) Plus DNS, screenshots, etc
No
wait
Is observation period over?
Sufficient evidence to continue?
No
Yes
Submit report to registrar / registry
Yes
Sufficient evidence to report?
No

Figure 1. Flowchart summarizing the evidence discovery and decision making process.

For the DNSTICR project, the final step of deciding if the domain should be reported has been taken by a human operator, although in the majority of cases the decision is fairly clear cut.

3 Statistics

Since the beginning of 2020 through to August 2021 we have seen 323,956 domains using one or more terms from our filter list. After the spike in registrations in April of 2020 we saw a clear reduction in the daily numbers as shown in Figure 2.

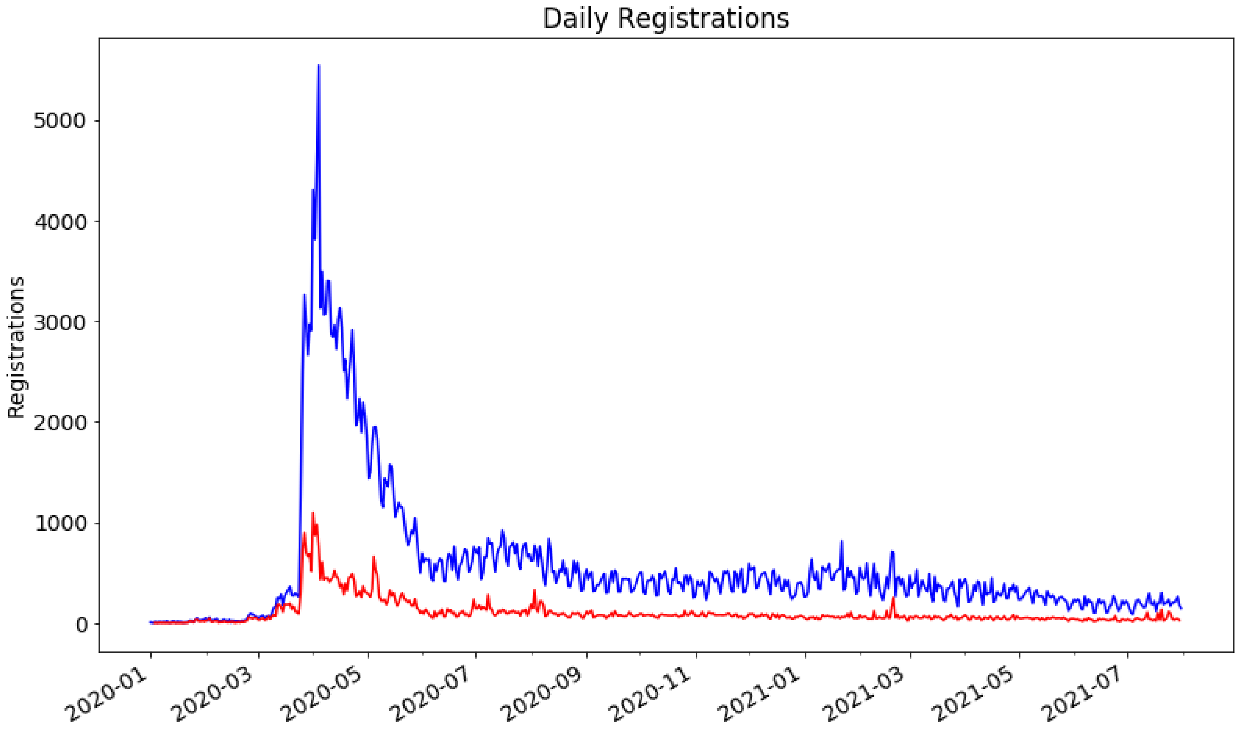


Figure 2. Registrations per day, new registrations that match a term in our filter list are shown in blue. The red line shows the domains for which we found one or more reports from reputation providers.

Through April of 2020 we identified 88,537 filtered domains, an average of 2,951 per day with a peak of 5,543 on the 4th April. This fell sharply to 50,389 domains in May and 23,952 in June. Since then the curve has largely flattened with just 5,979 registrations in July 2021 (compared with 26,652 in July 2020). These registrations can be seen in the blue line of Figure 2.

Of the 210,939 domains that were identified as being related to the pandemic from May 2020 to August 2021 the majority were benign or had a neutral or no reputation (*i.e.,* they did not resolve at all, were parked, or had no real content). In total 12,860 (6.1%) domains were able to be resolved and had one or more reports from domain reputation providers associated with them. The registration dates of these domains form the red line in Figure 2. This set includes domains labeled by a single, low-confidence, report. If we limit to high-confidence or multiple independent pieces of evidence we are left with 3,791 domains (1.8% of our filtered registrations) being flagged. As noted above, we search for evidence on the day we first see a registration and then every week over the next four weeks. Therefore, there is a chance that we miss the active period when a domain resolves and has evidence of malicious content either because it happens after our window or because the threat is dealt with before we revisit it.

The difference between the two lines in Figure 2 shows how many domains might be described as “suspicious” versus how many have any evidence of misuse.

There are, of course, several reasons why we will miss some malicious domains. We are going straight to the domains in question and so we do not see how the miscreants are directing people to their site. We can suppose that email or SMS is a common first contact, and that in some instances these messages will contain a full URL rather than a plain domain name.

Therefore, it is likely that we are missing some portion of the picture, although we do try to discover bad URLs under the domain name where possible.

We also see pages which redirect to a legitimate site or show little or no content, in some cases these will be targeted sites which only serve their malicious content to certain geographic regions or sometimes only to mobile devices. While we have had independent confirmation of a few cases of this happening we have not been able to quantify the scale of this behavior.

According to the scope of the DNSTICR project we are not looking into fraudulent or scam sites, so where we see for example facemasks or hand sanitizers for sale, we do not investigate further to see if any ordered products exist and are genuine.

Finally, it should also be noted that many of the terms we use, like “virus,” “mask,” or “payment” have uses and connotations not linked to COVID-19. These terms still match new registrations and these registrations may have evidence of malicious use (and so are counted above). We do not send reports for these domains as they are not related to the pandemic in any way and so they are not in the scope of the project.

The volumes of abuse registrations that we have seen through this project are somewhat lower than some other published figures. This difference is probably due to a number of factors, for example our reporting of unique registrations and not unique URLs or attacks. It is also the case, as has been noted above, that we are only looking at new registrations in generic top-level domains (gTLDs) not, for example, web traffic or SSL certificates. Other inputs or reputation sources may give very different results. We also have a narrower definition of abuse which does not include spam, fraud, or scam sites.

Our figure of 1.8% to 6.1% of COVID-19 related domains being malicious is consistent with a similar study which found 2.6% of the domains they examined having two or more positive detections on VirusTotal. That study also found that 85.3% of their domains had the COVID-19-related term at the second level as opposed to at a lower level (3rd level and below, *i.e.,* subdomains). This hopefully indicates that by only looking at registrations in TLD zone files, and not seeing any subdomains being used, we are not missing too many malicious sites.

4 Trends

Within the data we are collecting we see other, more subtle trends playing out.
Figure 3 shows the top matching terms from our list (any term which contributes over 1% of the total is shown).

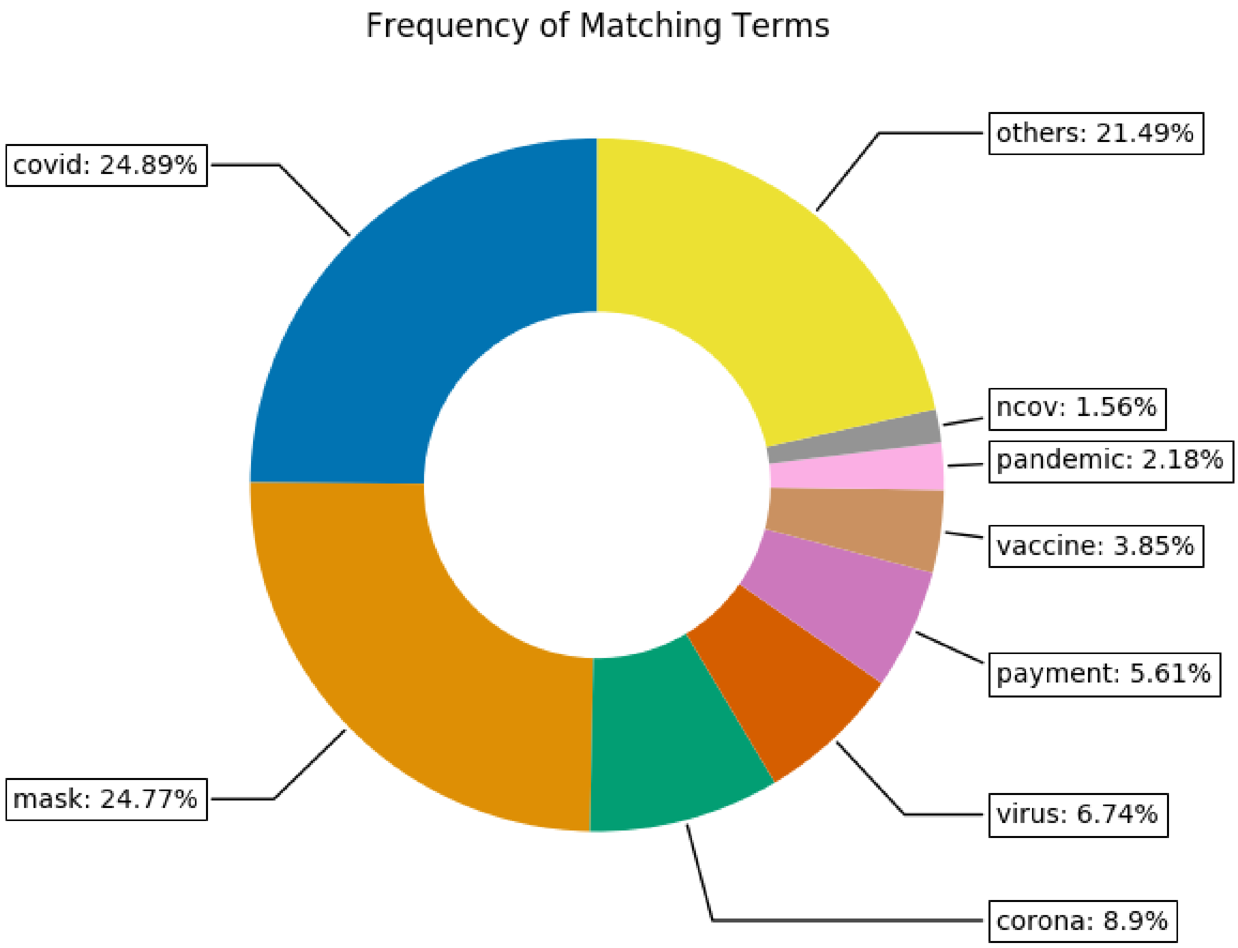


Figure 3. Frequency of matching terms. Terms matching all new registrations from January 2020 through to August 2021. This does not reflect the frequencies within the set of terms that had some evidence of malicious use, for those see Figure 7.

We can also see how these proportions changed over time. Figure 4 shows the top 10 matching terms as daily percentages since the beginning of March 2020 (note that registration levels vary greatly over this timescale).

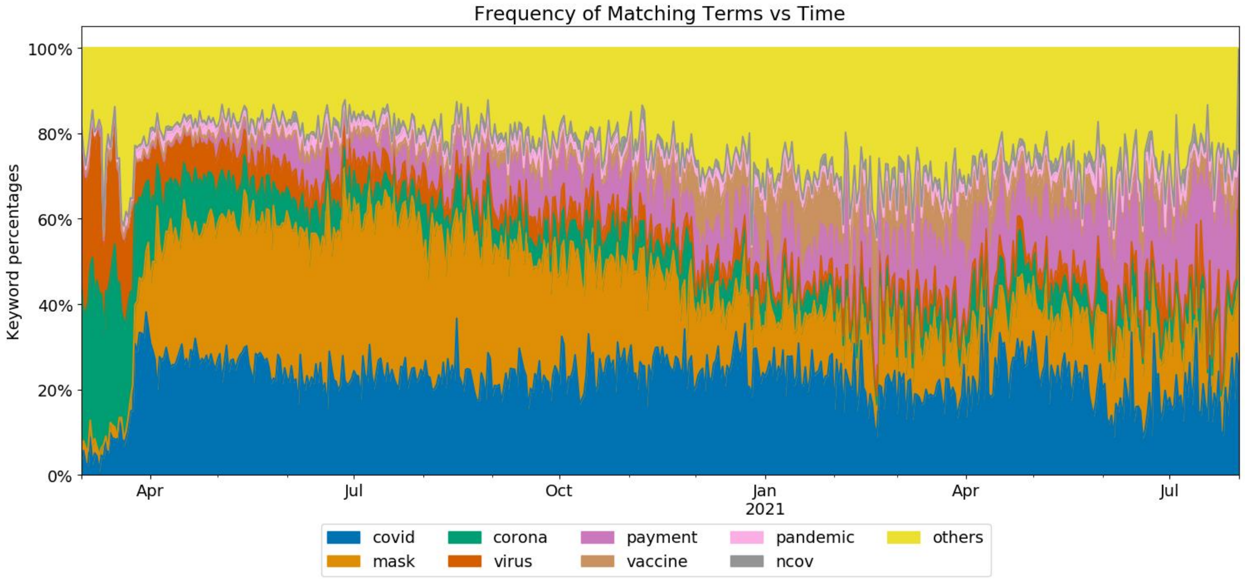


Figure 4. Frequency of matching terms versus the time period of March 2020 to the end of July 2021.

Note how, since late March, the term “covid” has maintained a fairly steady proportion of matches, while the term “corona” has seen a relative decline after an initial spike.
Harder to see in Figure 4, but shown in Figure 5, is how contributions from the term “vaccine” have evolved over time. We can see that the volume of registrations matching “vaccine” started to grow in November 2020, dipped around Christmas and peaked (with the odd exceptional day) in mid-January 2021.

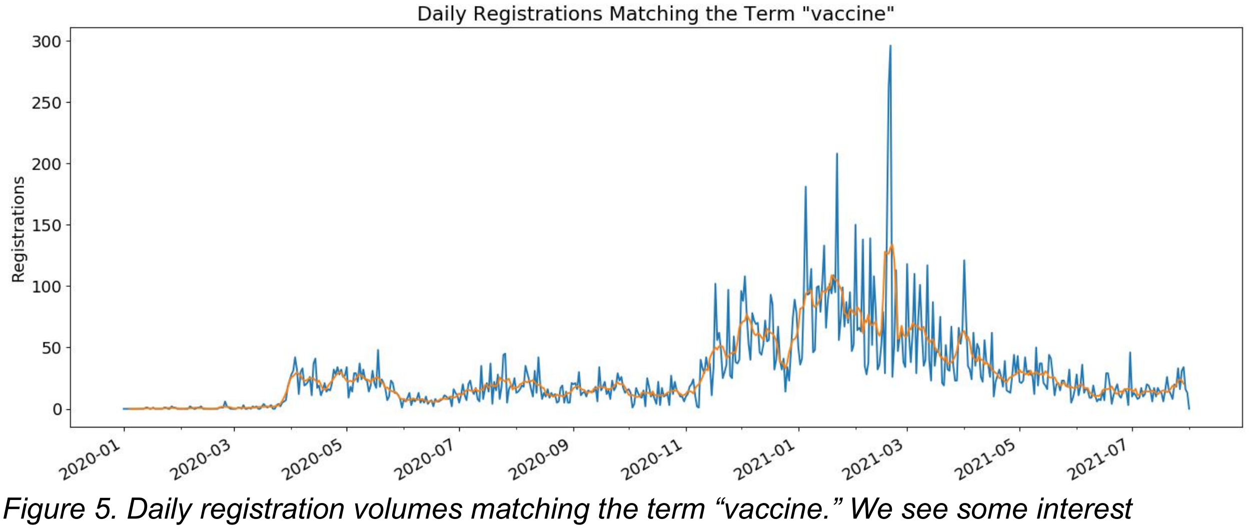


Figure 5. Daily registration volumes matching the term “vaccine.” We see some interest throughout our data, but in particular around the end of 2020 and the first quarter of 2021. The orange line shows a 7-day rolling average.

These trends reflect what we see from other observations of how the language of the pandemic was changing over time, for example Google Trends shows similar patterns, see Figure 6.

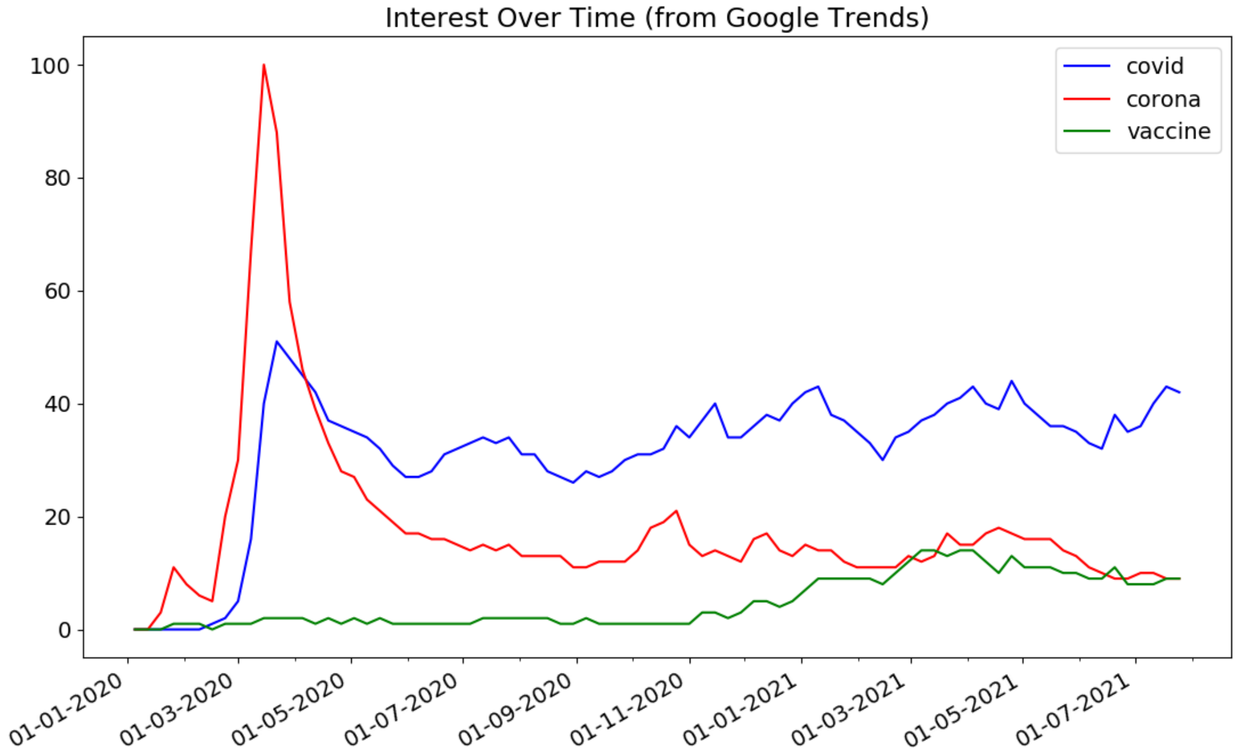


Figure 6. Interest over time as seen by Google Trends. Looking at just three of our terms, measured weekly since the start of 2020. One hundred indicates peak search interest across the set.

If we look at trends in just those domains which had at least one third-party report we see a similar picture but with some small differences. For example if we look at the overall prevalence of terms, shown in Figure 7, we see that “corona” and “virus” have overtaken “covid” and “mask.” The other terms all occupy the same positions and no new terms contribute more than 1% of the total number of domains found. This potentially reflects the volumes matching those terms at a time when a larger number of registrations were being flagged.

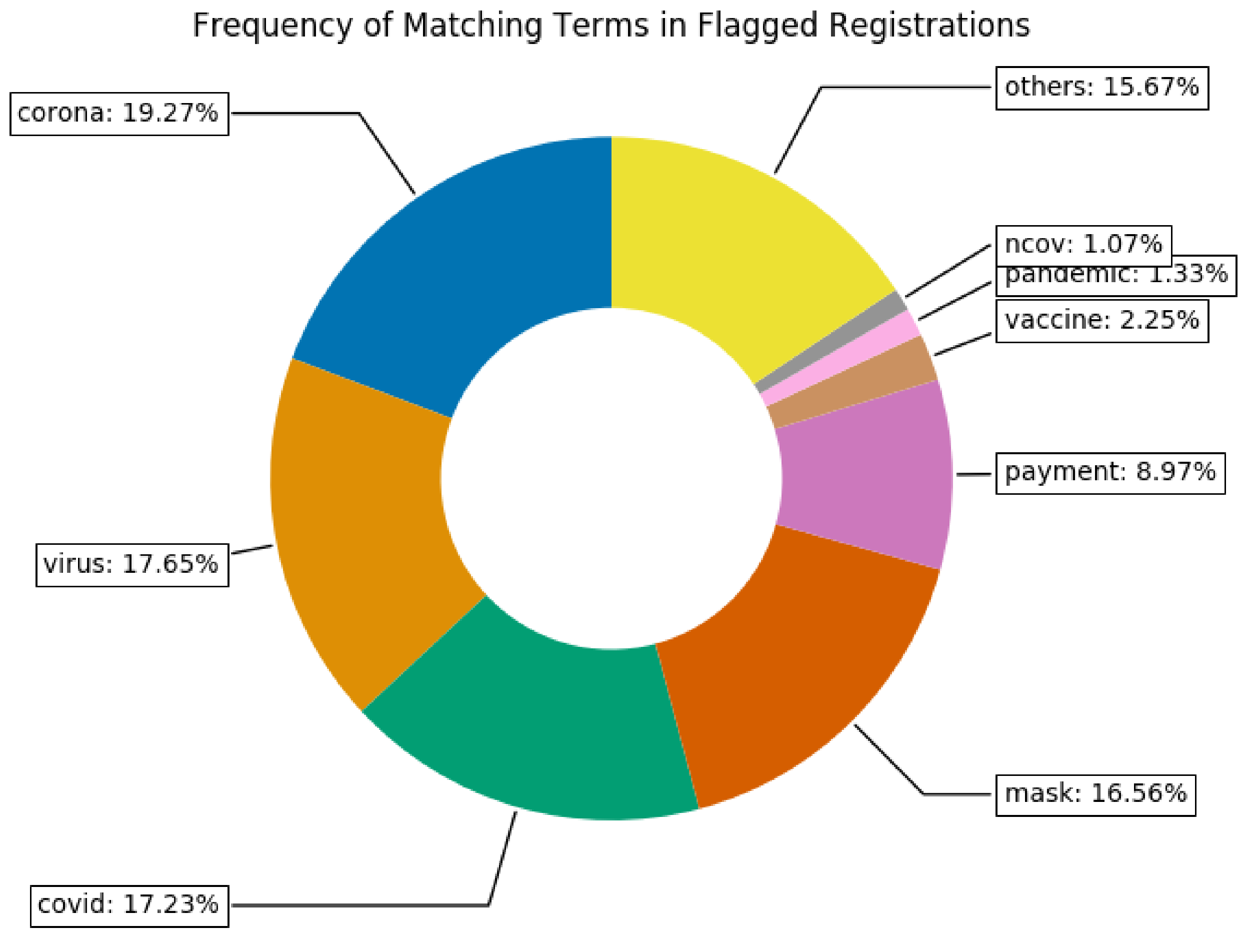


Figure 7. Frequency of matching terms seen in flagged registrations. That is, registrations which appeared on one or more third-party reputation lists. Compare with Figure 3 which shows the frequencies among the parent set of all COVID-19 related new registrations.

We can again see how the prevalence of these terms changes over the course of our observations to date, as shown in Figure 8.

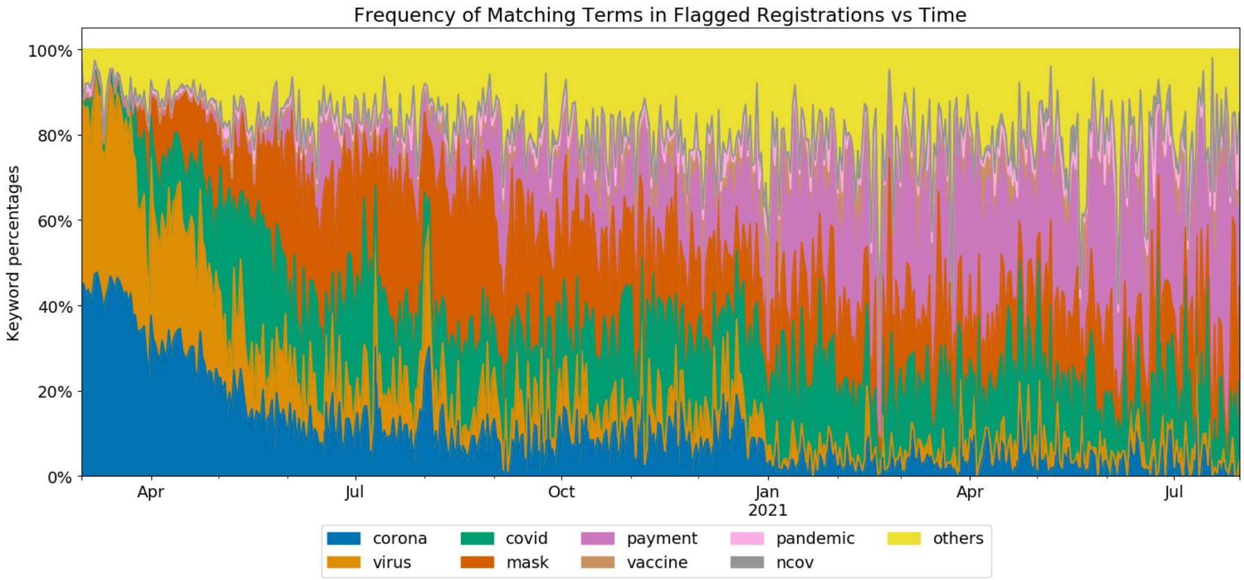


Figure 8. Frequency of matching terms in flagged registrations versus time. For comparison see Figure 4 which shows the same calculation in the parent set.

As in Figure 4, we see the frequency of “corona” falling in the first few months while “covid” remains at a fairly constant level.

We can again pick out certain keywords, like before (Figure 5) we can look at the term “vaccine”:

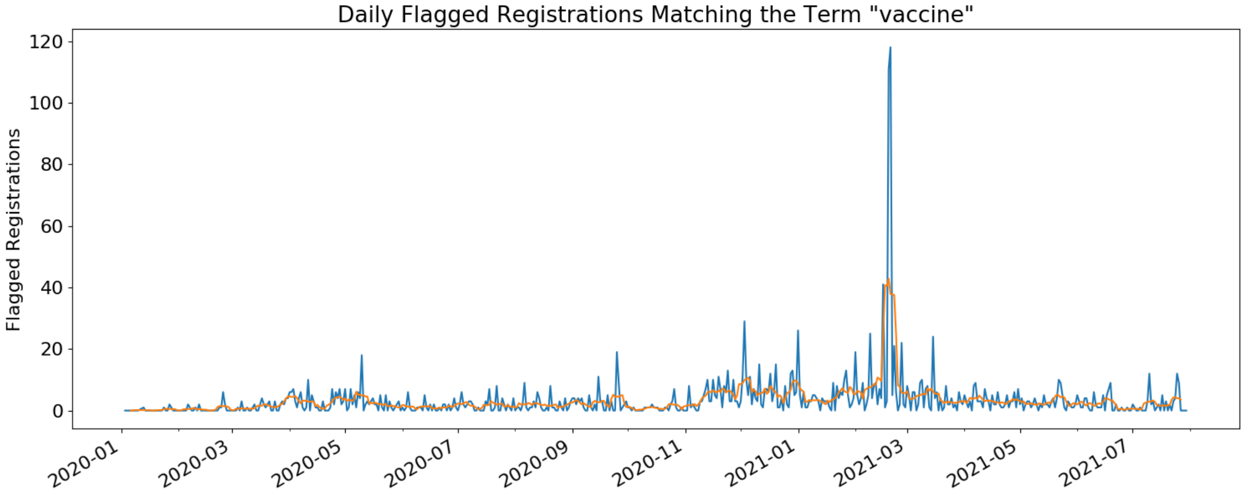


Figure 9. Daily flagged registration volumes matching the term “vaccine,” are compared with full registration volumes shown in Figure 5. The orange line shows a 7-day rolling average.

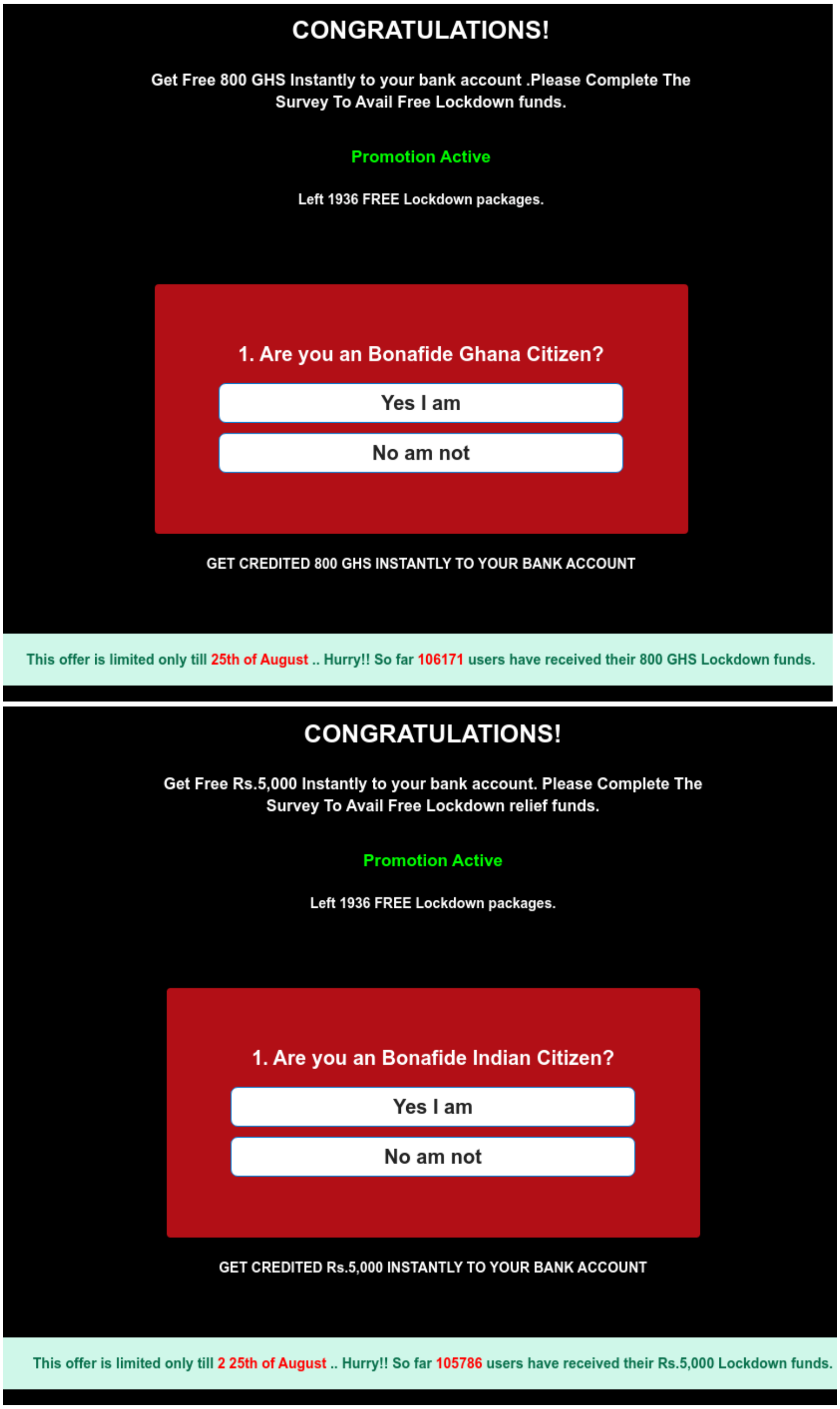
The large spike in Figure 9 occurs on the 19th and 20th February 2021 where we see around 110 registrations each day that begin “myvaccine” followed by a U.S. state name, a close misspelling of a state name, or a two-letter state abbreviation. It is likely that these domains were flagged by a third-party using registration patterns as part of their detection mechanism, when we looked closer though, we only found parked pages and no evidence of malicious behavior.

5 Examples

The statistics we show above tell a certain story, but do not give a feel for the types of campaigns we have observed. Over the time we have been investigating domains, a number of trends and campaigns have been seen. We do not claim to see everything and may well be missing important events, but still, within our data there are some interesting observations to be made.

When we do find pages related to phishing, what do they look like? Basically, they appear like any sort of login page where there is some value (most likely financial) in harvesting Internet users' credentials. Often pages will look exactly like a banking or ecommerce login page (maybe with an additional message about COVID measures). Many of these campaigns can be seen operating before the pandemic albeit with more generic looking domains.

Another common tactic is to offer a financial payment of some description, possibly in the guise of a COVID or lockdown stimulus payment. These are typically tailored to a specific country, although we do see the reuse of some templates. Figure 10 shows two examples from a set of near-identical pages where only the currency and country have been changed. In each of these cases the user was eventually led to a (fake) Facebook login page that would harvest any credentials used.



CONGRATULATIONS!

Get Free 800 GHS Instantly to your bank account .Please Complete The Survey To Avail Free Lockdown funds.

Promotion Active

Left 1936 FREE Lockdown packages.

1. Are you an Bonafide Ghana Citizen?

GET CREDITED 800 GHS INSTANTLY TO YOUR BANK ACCOUNT

This offer is limited only till 25th of August .. Hurry!! So far 106171 users have received their 800 GHS Lockdown funds.

CONGRATULATIONS!

Get Free Rs.5,000 Instantly to your bank account. Please Complete The Survey To Avail Free Lockdown relief funds.

Promotion Active

Left 1936 FREE Lockdown packages.

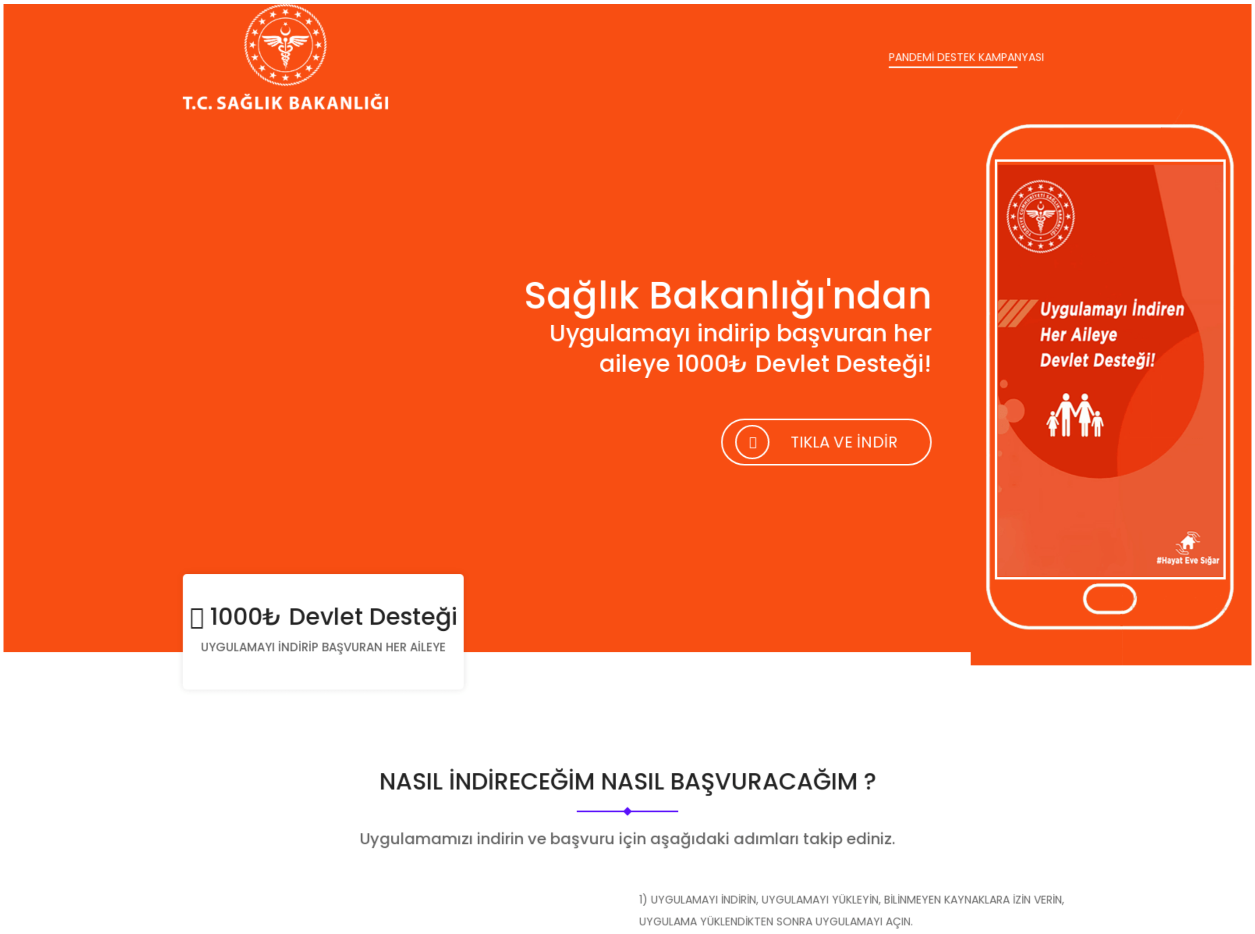
1. Are you an Bonafide Indian Citizen?

GET CREDITED Rs.5,000 INSTANTLY TO YOUR BANK ACCOUNT

This offer is limited only till 2 25th of August .. Hurry!! So far 105786 users have received their Rs.5,000 Lockdown funds.

Figure 10. The country and currency changes, but everything else about this template status is the same.

A third common tactic we observed was to appear like an official government campaign for example as seen in Figure 11, we saw many variations around this theme.



T.C. SAĞLIK BAKANLIĞI

PANDEMİ DESTEK KAMPANYASI

Sağlık Bakanlığı'ndan
Uygulamayı indirip başvuran her aileye 1000₺ Devlet Desteği!

TIKLA VE İNDİR

Uygulamayı İndiren Her Aileye Devlet Desteği!

#Hayat Eve Sığar

1000₺ Devlet Desteği

UYGULAMAYI İNDİRİP BAŞVURAN HER AİLEYE

NASIL İNDİRECEĞİM NASIL BAŞVURACAĞIM ?

Uygulamamızı indirin ve başvuru için aşağıdaki adımları takip ediniz.

1) UYGULAMAYI İNDİRİN, UYGULAMAYI YÜKLEYİN, BİLİNMEYEN KAYNAKLARA İZİN VERİN, UYGULAMA YÜKLENDİKTEN SONRA UYGULAMAYI AÇIN.

Figure 11. A webpage that appears to be from the Turkish Ministry of Health.

The targets in this specific case are Turkish nationals, with an offer of 1,000 Turkish Lira from the Turkish Ministry of Health. To complete their application the visitor is invited to “click and download,” however, the file linked to is detected as malware by many antivirus engines (around half of those on VirusTotal).

Another tactic which we have seen involves parcel delivery notices, either missed deliveries or more often referencing an outstanding payment required before delivery can take place. While not directly related to the pandemic, an increased volume of online shopping during lockdowns makes these schemes more likely to succeed. These sorts of scams initially ask for a small sum of money, but can eventually lead to much larger fraud as documented in various news stories such as the one here.

More recently we have been seeing a number of malicious domains posing as the U.K. National Health Service’s “COVID pass” application form. At the time of writing it is too early to tell if this will be a sustained trend or a short term phenomenon.

6 Conclusion

Over the past 18 months we have seen a surge, and fall, in new domain registrations that match a set of keywords related to the COVID-19 pandemic. While the majority of these registrations have not been observed to be harmful in any way, a minority have been identified as malicious.

Over the course of our investigation the tactics used by domain speculators and malicious actors have evolved, tracking the way the pandemic has impacted on the world.

Malicious campaigns have been seen along expected lines of offering incentives, often financial, or posing as a legitimate log-in page in order to steal credentials or deliver malware. The only difference is that in our set the “hook” used to lure victims in involves COVID-19 in some fashion.

DNSTICR is an ongoing project that continues to evolve to the everchanging COVID-19 pandemic. OCTO will continue to provide updates on this project to the ICANN community, security professionals, and Internet users.