Unraveling Threat Intelligence Through the Lens of Malicious URL Campaigns

Mahathir Almashor∗
mahathir.almashor@csiro.au
CSIRO Energy
Sydney, Australia

Ejaz Ahmed∗
ejaz.ahmed@data61.csiro.au
CSIRO Data61
Sydney, Australia

Benjamin Pick∗
benjamin.pickle@data61.csiro.au
CSIRO Data61
Sydney, Australia

Jason Xue∗
jason.xue@data61.csiro.au
CSIRO Data61
Sydney, Australia

Sharif Abuadbba∗
sharif.abuadbba@data61.csiro.au
CSIRO Data61
Sydney, Australia

Raj Gaire∗
raaj.gaire@gmail.com
CSIRO Data61
Sydney, Australia

Shuo Wang∗
shuo.wang@data61.csiro.au
CSIRO Data61
Sydney, Australia

Surya Nepal∗
surya.nepal@data61.csiro.au
CSIRO Data61
Sydney, Australia

SYNOPSIS

The daily deluge of alerts is a sombre reality for Security Operations Centre (SOC) personnel worldwide. Those on the front-lines of cybersecurity face the unenviable task of prioritising threats amongst a flood of URLs found within malicious communications. Timely detection of pertinent patterns within such URLs allows teams to deescalate threats. This has been traditionally filled with machine-learning log analysis and anomaly detection methods. Instead, we propose to analyse suspicious URLs from the perspective of malicious URL campaigns. By first grouping URLs within 311M records gathered from VirusTotal into 2.6M suspicious clusters, we thereafter discovered 77.8K malicious campaigns. From those, we found 9.9M unique URLs attributable to 18.3K multi-URL campaigns that had at least 1 URL flagged by a vendor within VirusTotal. Worryingly, our analysis shows that only 2.97% of such campaigns were detected by security vendors. We also confer insights on evasive tactics such as ever lengthier URLs and more diverse domain names, as well as case studies that expose other adversarial techniques. By characterising the campaigns driving these URL alerts, we hope to expose current threat trends, and arm the community with greater threat intelligence.

1 INTRODUCTION

Timely and decisive threat intelligence is a vital imperative for cybersecurity personnel defending their organisations. Within their respective Security Operations Centres (SOC) [22, 24], human teams are inundated with alerts for every potential incursion, and are often under great pressure to rapidly recognize and react to high priority threats [3, 43, 52]. Security Information and Event Management (SIEM) [13, 17] systems present in most SOCs collect and compile real-time events from an organisation’s end-users, applications and devices [35]. These are then surfaced to SOC analysts to sift through and verify, with good threat intelligence playing an outsize role in identification and prioritisation efforts [50]. Machine learning based logs analysis [8, 38], data aggregation and inference [33], and anomaly detection [2] all aim to lower the burden for human operators in such settings.

Aforementioned approaches can be improved since we cannot simply rely on a huge swath of Uniform Resource Locators (URLs) classified into benign or malicious by deep learning. This is because such approaches rely on training on past data, and are ill-equipped to handle any permutations or evolution of threats. For learning based detectors, threat intelligence such as this could help tease out features and connections amongst disparate URLs. To this end, we propose an alternative that focuses on bettering threat intelligence through the lens of malicious URL campaigns. Specifically, we group the URLs using the hash of the content they’re pointing to, and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:
Mahathir Almashor, Ejaz Ahmed, Benjamin Pick, Jason Xue, Sharif Abuadbba, Raj Gaire, Shuo Wang, Seyit Camtepe, and Surya Nepal. 2023. Unraveling Threat Intelligence Through the Lens of Malicious URL Campaigns. In Asian Internet Engineering Conference (AINTEC ’23), December 12–14, 2023, Hanoi, Vietnam. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3630590.3630600

∗ Commonwealth Science and Industrial Research Organisation (CSIRO)
use that to tease out the patterns in seemingly disparate groups of URLs. The reasoning here is that the campaigns often use the same payload (e.g., a phishing site, or virus-laden file) that resolve to the same hash.

This is pertinent as SOC teams are inundated with suspicious URLs, be it from communication channels (e.g., within a phishing email), or from application and network events (e.g., a compromised device phoning home). Malevolent actors exploit this weakness by incorporating URLs into multi-stage strategies, with a reliance on waves of diverse URLs to maximise reach and evasion. We witness the recent surge of attacks based on the COVID-19 pandemic [7], where attackers continued their use of carefully crafted URLs to deceive users [5, 12]. This leads to the divulging of credentials [32], or installation of malware [49].

Such concerted use of URLs are what we define as malicious campaigns. That is, a set of URLs that all point to the same content, where at least one URL flagged as malicious. Regardless of the vector, nature and time of security threats, attacks can be devolved down to the URL. This supplies the impetus for us to cluster and characterise seemingly distinct URLs as a motivated whole, via a practical yet effective mechanism: the hash of content pointed to the URLs themselves.

We show that we can trivially recognise suspicious collections of URLs at scale. To replicate the torrent of URLs faced by SOC teams, we gathered 311M URL submissions stretched over 2 months from VirusTotal [55]. This serves as a strong parallel for SOC issues, as our URLs originate from varying sources, which resembles how SOCs are alerted of URLs via multiple SIEM events. Further, our data has labels from security vendors of varying origins and quality, not unlike the diverse labelling seen in SIEM alerts. Thus, by applying content-hash grouping to this data, 2.6M suspicious clusters were exposed which led to the following contributions:

- The detection of 77,810 confirmed malicious campaigns via the reported content hashes within each submission.
- 6 pertinent findings relating to characteristics for 59,450 single-URL and 18,360 multi-URL malicious campaigns.
- 3 selected case studies confirming the impact of campaigns and the shifting tactics used by attackers to evade defences.

The contributions above have allowed us to quantify and articulate the influence of URL campaigns on threat intelligence efforts, which aids SOC personnel in prioritising the most serious threats and thus, amplifies their proficiency.

1.1 Ethical Considerations

This work does not raise any ethical issues. The obtained dataset purports to study Internet measurement in good faith, and is secured privately with access granted only to the authors’ affiliations. The dataset does not include any personally identifiable information, with all examples in this work anonymised so as to prevent linking sensitive activities to specific organizations or users.

2 BACKGROUND & RELATED WORK

Complexity and persistence of cyber-attacks surge to new heights at scale, and threat intelligence plays an ever vital role in helping human security teams manage an otherwise overwhelming volume of information [1]. It is a culmination of the old adage of “knowing your enemy”, where the emphasis has shifted to proactive defence instead of purely reactive approaches [21]. We know that adversaries eschew isolated attacks in favor of coordinated campaigns [29], which only underlines the gravity of the work herein.

Given that campaigns are central to threat intelligence, the research gap we strive to address is the measurement of orchestrated malicious use of URLs. Colloquially known as links, URLs are simply pointers to Internet content [10]. Their simplicity and ubiquity invites abuse within multiple contexts, from social media [26], to text messages [42], and compromised devices [60]. For SOC teams defending their systems, URLs are key Indicators of Compromise (IOC) that feed into human decisions [59]. For end-users, their deceptive use [4, 40] is best exemplified in the second example below:

| Field | Examples |
|-------|----------|
| url   | http://mail.get-apple.com/support/index.php |
| content hash | f5686646fe7ae7c7...dc2dc840230f4215e |
| positives | \(0 \leq n < N\), where \(N\) is max vendors |

The scheme for most URLs remains http, though rising malicious use of https is noted in the prior art [16]. Sub-domains are fully within the purview of attackers, and is oft-used to disguise URLs [40]. Domain rights are restricted depending on the suffix in question. For example, Apple can register against .com.au and .com.sg, via resellers entrusted by their respective governing entities: auDA [6] and SGNIC [47]. Such domain authorities have processes in place that verifies the registrant (e.g., Apple) and mitigates against malicious squatting of domains. It must be noted that management of suffixes (also known as top level domains) can vary depending on the authority in question, leading to varying levels of security rigorous. Path refers to the exact location of a page, file or asset (i.e., the actual content) on the addressed server.

2.1 VirusTotal & Vendor Labelling

Prior focus on harmful URLs has often treated them as individual pieces, with works ranging from detection at the domain name level [15, 20, 34, 53, 54, 57], deep-learning classification of URL strings [14, 38, 45], and the binaries that the URLs point to [27, 28, 60]. Given the historical scale of the issue [31], streams of suspicious URLs are sent to online security vendors, with each employing customised engines to classify malicious URLs with varying levels of success [39].

Building atop of such services, the VirusTotal (VT) platform acts as a results aggregator [51]. Essentially, it allows convenient querying of suspicious URLs across multiple vendors. While this offers simplicity, it is noted in prior works that vendor labels do not always agree, and that VT based defensive mechanisms are not sufficient [37, 44, 48, 58]. As such, we carefully distill the labels within our dataset, and reinforce them with results from Google Safe Browsing (GSB) [18].
3 APPROACH

In this section, we detail each step of our content-hash based clustering of malicious campaigns, the secondary verification that was performed, and finally, show the potential impacts of content-hash detection.

3.1 Dataset Description

Our goal was to approximate the constant stream of URLs surfaced by SIEM systems. To that end, we used 311,588,127 URL submissions over a 2-month duration (2019-12-02 to 2020-01-31) gathered with kind assistance from VirusTotal. Table 1 shows selected metadata fields within each submission. Here, we focus on the "content hash" and "URL" fields, with the former being the computed hash of the content to which the latter points. That is, both bad.com/mal.exe and new.com/run.exe point to the same malicious executable, this would result in the same content hash.

We note that 11.39% of submissions (over 35.4M) had missing content hashes. These exhibited no clear patterns and were uniformly spread, thus ruling out periods of accidental corruption. Further, 77M unique URLs were found, pointing to a naive mean rate of re-submission of 4.044 per URL. Of these, 24.5M URLs (31.9%) were flagged by an average of 71.9 vendors, with a majority (94.18%) checked by 72.

The dataset approximates the temporal characteristics of SIEM alerts for an SOC team within a typical target organisation. We see this in Figure 1, where daily scan volumes were divided into three 8-hour chunks starting midnight, 8am and 4pm. With no timezone data, all times were taken to be UTC. This would be analogous to the constant barrage of threats faced by the SOC team, both internally reported and as adversaries prod their defences. Note that outliers were observed in daily volumes from December 2nd to 5th, but after fitting a linear model to our weekly averages, we found a slope of -0.0290, which points to a steady rate of submissions.

3.2 Content Hash Clustering

Once submitted, the content pointed to the URL is retrieved and a SHA-256 hash function is applied. Figure 2 illustrates how webpage 1_machine.com/login and new_machine.com/run.exe point to the same malicious executable, this is simply the sum of all Positives divided by the number of submissions in each cluster.

To further our analysis, we needed to confirm the malicious nature of the 2.6M flagged clusters. We cannot rely on original vendor labels as there is contention as to the correct threshold to apply. [11, 19, 34, 41, 46, 54, 60] have set their thresholds for maliciousness to $t = 1$, with $t = 2$ being used in the previous works [38, 48, 56], and even $t = 5$ in the prior work [25]. Further, there is an opportunity to compare existing labels against present day data. Thus, to check if clusters were indeed malicious, we implemented secondary verification via GSB, which is noted to be the largest block-list [9]. We then expand our definition:

- **Malicious Campaigns**: Flagged clusters that are further verified by GSB as containing at least one malicious URL.
3.5 Initial Findings

Using this direct approach, we discovered 77,810 confirmed malicious campaigns, which represented 2.97% of the 2.6M aforementioned flagged clusters. These campaigns had at least one URL within them marked by GSB as either malicious (MALWARE), deceptive (SOCIAL_ENGINEERING), or both.

Finding 1: With only 2.97% campaigns detected, we see the limitations of current block-list approaches. On average, the mean positive score for these campaigns stands at 6.2 (∼3.05). Using this as a threshold, 605,323 other flagged clusters with higher mean positive scores would have eluded our secondary verification. That is, if assuming all 2.6M flagged clusters were malicious, GSB would have missed 605,323 campaigns containing 2,034,777 malicious URLs.

Finding 2: Performance of GSB secondary verification drops as campaigns use more URLs. We observed that detection by GSB degrades as campaigns grow larger, with a 13.27% rate for campaigns that deploy ≥ 100 URLs as compared to 74.32% for ones with ≤ 100 URLs. The largest campaigns exhibit even lower rates at 2.46%, which may be attributed to the limited life-spans of malicious URLs. This agrees with the work in [9], which reports that fewer URLs remained in block-lists as time progresses. This also hints at the ephemeral nature of such URLs, and on the fact permanent banning policies are not enforced, although it was also noted that numerous URLs tended to reappear after removal. Similarly, vendors struggled to detect URLs as campaign sizes grow. We observed that on average, six vendors detecting campaigns of size 100 and below, and only four detecting larger campaigns (size over 100).

3.6 Broader Impact of Content Hashes

A consequence of our work is perhaps that the detection of threats within SIEM alerts can be improved if SOC teams can factor in content-hashes at the triage stage. Blocking singular URLs works on past but not future threats. Using content-hashes, previously unseen threats could be rapidly identified in the absence of other obvious malicious connections.

We see evidence of this with the small malicious campaign below, where one URL (everyday-vouchers.com) remained unflagged for six submissions before being flagged by GSB on 2020-01-02. Soon after, another URL (sweepstakehunter.com) bearing the same content hash was observed as remaining unflagged for 5 submissions to the end of the collected data.

| Malicious Campaign ID | Everyday-Vouchers.com, Sweepstake-Hunter.com |
|-----------------------|---------------------------------------------|
| Mean Positive Score   | 0.08333 (across 12 submissions)             |

One year later, both URLs were marked as malicious by GSB in our secondary checks. In this case, a SIEM threat analyser based on content-hash would have raised the priority alert for the second URL as soon as it had appeared. It is noted that attackers could automate the mutation of malicious content for each end-user visit. This would evade content-hash detection as presumably, each URL results in a unique hash. Accordingly, teams can fall back on methods such as linking IP addresses and server signatures, or on HTML comparison methods such as in [30]. Nonetheless, content hashes are still useful for clustering URL alerts before more advanced mechanisms are brought to bear.

4 MEASURING MALICIOUS CAMPAIGNS

Preceding sections may have only affirmed an instinctual knowledge within SOC teams: that the daily torrent of SIEM alerts about suspicious URLs are rarely isolated events. We thus dive deeper into our discovered malicious campaigns.

Finding 3: Vast majority of malicious URLs are from campaigns employing multiple URLs. Table 2 shows the general details of confirmed campaigns, which we have broadly designated into campaigns that have either a single unique URL, or multiple unique URLs. A majority (76.4%) of the found campaigns consists of single URLs (i.e., the same URL is repeatedly submitted) but these only contributed 59,450 (0.6%) unique URLs and 276,801 (0.73%) submissions. The 18,360 multi-URL campaigns provided the bulk of URLs and submissions, lending credence to our hypothesis that masses of URLs should be treated as coherent campaigns.

Finding 4: Campaigns favor the use of longer URLs. The average URL lengths across campaigns (i.e., mean of means) stands at 64.29 characters. There is a standard deviation of 85.76 indicating a high degree of variance which we see in Figure 4 with some campaigns

![Figure 3: Cluster distribution by submission counts](image-url)
on the right extreme employing URLs with tens of thousands of characters. 50.89% of campaigns have average URL lengths of between 38.0 (1st quartile) and 73.0 (3rd quartile). This is likely due to attackers relying on increasingly longer URLs to evade detection. As awareness increases in end-users, they grow to distrust shorter URLs pointing to unknown locations [4] or short URLs with added terms in their domain names [40].

4.1 Campaign Metrics

Next, to systematically quantify malicious campaign behaviours, we define and calculate the following metrics:

- **Campaign size**: Total number of URLs deployed.
- **Source distribution**: Ratio of the campaign size to the number of submissions. A 100% source distribution indicates that a different URL is used for every submission. This quantifies the rate at which attackers attempt different URLs, and the URL diversity within a campaign.
- **Campaign footprint**: Here, we consider metrics from both attackers (via a count of unique URLs in a campaign), and victims (number of times URLs were reported). The following expression is used to compute the footprint: \( f_p = \mu U + (1 - \mu)S \), where \( U \) represents the count of URLs in a campaign and \( S \) denotes the number of times these URLs were submitted. \( \mu \) represents the weight assigned to each metric, with equal weights of 0.5 set for both sides and \( \mu \) able to prioritize certain metrics.
- **Domain diversity**: Ratio of unique domains to the number of URLs in a campaign, where a 100% score indicates that each URL has used a unique domain. This metric aids us to understand if a campaign is targeting specific global brands, or if it distributes attacks across a range of users. Conversely, low domain diversity can suggest a highly targeted campaign towards an organization.
- **Sub-domain diversity**: Same as domain diversity above but for sub-domains within the URLs of a campaign.
- **Detection**: Here, we evaluate detection rates using two metrics: GSB and the original vendors within the data. A 100% GSB detection means that all of a campaign’s URLs were flagged by GSB. For vendors, we report the mean number of vendors who flagged a campaign’s URLs. This helps us understand the general detection rate of URLs.

4.2 Campaign Insights

**Finding 5**: Most multi-URL campaigns deploy different URLs in their attacks. The average source distribution for all campaigns stands at 61.58%. We observed that 6,160 campaigns with an average campaign size of 146 URLs (footprint of 160) had source distribution of 98.75%. Moreover, 5,498 campaigns had source distribution of 100%. These high source distributions show that campaigns are frequently creating new attack URLs, and implies that traditional allow/block-lists will not effectively detect URLs in many campaigns.

We also observed that for 1,040 larger campaigns with campaign size greater than 100 URLs, the source distribution reduced to 35.95%. For such campaigns, the average campaign size, footprint, and number of reported submissions were 5,132, 19,682, and 34,233, respectively. Moreover, we observed that the largest campaign with a campaign size of 1,391,080 URLs (footprint at 4,977,600, and submissions at 8,584,120) had a source distribution of 16.21%.

These findings indicate that the source distribution of a campaign decreases with increases in campaign size. This may be attributed to the fact that as the number of URLs increases, the corresponding reports from users (i.e., submissions) increases non-linearly. This indicates that larger campaigns may be using domain generation algorithms (DGAs) to generate a swathe of domain names for both targeting, and command and control. This large number makes it arduous for vendors to take down URLs from such campaigns, as infected computers will attempt to contact some of these domain names daily to receive updates or commands.

**Finding 6**: We discovered several campaigns impersonating well-known global brands. Given the finding above, we found campaigns using multitudes of URLs to mislead victims by including company or product names. The example below shows a campaign that targeted Apple with 4,081 unique URLs. It used a combination of 9 sub-domains, 12 domains, and 7 suffixes, with the ratio of unique domains to URLs standing at 0.29%, using the following variations:

- **Sub-Domains**: www.apple.com, www.icloud.com, www.apple.icloud.com, apple.apple.com, icloud.com, mail
- **Domains**: icloud-com, online-support, findmy, get-apple.com-support, map-apple.id-info, www.wvvw.icloud, sign-in, map-log.com-fml, viewlocation-icloud
- **Suffixes**: us, in, support, live, review, com, mobi

We see the precise targeting of not only the company name, but various associated products. The particular content hash for this campaign was submitted 104,311 times, and we observed that an average 11 vendors flagged URLs within this campaign, compared to the average 3.4973 flags across all 311M submissions. Surprisingly, we also observed that GSB only detected two of the URLs.

**Finding 7**: Campaigns had an average domain and sub-domain diversity of 66.57% and 21.45%, respectively, with specific targeted organizations. We observed higher diversity in domain names compared to that of sub-domains. This is because attackers are limited to certain guidelines when registering domain names whereas sub-domains can be chosen freely. Due to this flexibility, campaigns targeting particular organizations can use sub-domain names specific to the target organization, such as in the targeted phishing above. We observed that a small number of campaigns (153) had domain and sub-domain diversities of 2.29% and 1.61%, respectively.
This indicates that attackers used fixed sets of domains and subdomains targeting a few victim brands.

Finding 8: 98.93% (1,572/1,589) of URLs within a known malware campaign used transport layer protocol (TLS). Continuing the trend of diverse malicious URLs, we found a campaign aimed at installing malware into victim machines, with a size of 1,589 URLs and a source distribution of 16.57%. The campaign used 261 unique domain names and a total of 1,589 URLs which were submitted 9,589 times. 1,175 of the URLs had the domainubar-pro4.ru, which was used specifically for malware distribution [23]. This finding is in line with [36], which reports that up to 25% of malware rely on TLS. Encryption is a potent evasive tactic that obfuscates code and/or the path component of URLs, and we note that these URLs lure users to download malware with file extensions such as: exe (19), js (28), zip (7), and rar (3).

We also discovered that the campaign’s URLs pointed to malicious torrent files. So, instead of typical links to a web-page, the URLs forwards the victim to a torrent protocol which allows malware to be downloaded via numerous torrent peers rather than a single server. This enables strong persistence of malicious content even when trackers are unavailable. We observed 46 (2.89%) URLs with such capability, with very low detection rates for GSB and VT vendors. From the 1,589 unique URLs, only two were detected by GSB and on average, only one vendor flagged the URLs within.

4.3 Fileless Malware

Campaigns also exhibit sophisticated tactics that leverage native pre-installed system tools (colloquially named “living off the land”). Such fileless malware attacks forego the need for payloads that can trigger existing disk scanning tools.

Finding 9: We observed 23 campaigns that leverage a system’s pre-installed software for compromise. The URLs within such campaigns are representative of a three-stage process: First, the infection typically starts with phishing emails rigged with deceptive URLs. Second, a malicious payload is downloaded using a command-line tool (e.g., cmd.exe). The Windows PowerShell tool can also be invoked with heavily obfuscated code to evade analysis. The payload is kept in-memory rather than on-disk, where it is finally executed, whereupon its plugins are loaded and a communication loop is entered that fetches tasks from command servers. The example below shows malicious code embedded within a URL that downloads and executes bad.exe via PowerShell:

```plaintext
http://xxx.xxx.xxx[x].xxx/public/invokefunction&
function=call_user_func_array&
vars[0]=system&
vars[1]=
[
]<cmd.exe>/c\powershell\(new-object System.NetWebClient).DownloadFile
('www.bad.com/download.exe',\SystemRoot\)
\Temp\bad.exe\]start\SystemRoot\Temp\bad.exe
```

Upon further analysis, we found other legitimate system tools abused for malicious tasks, including:

- winmpvse.exe
- svchost.exe
- conhost.exe
- lam.exe
- system.exe
- lssas.exe
- winlogon.exe
- spoosl.exe
- csrss.exe
- securityhealthservice.exe
- services.exe
- wininit.exe
- cmd.exe
- explorer.exe
- ssms.exe
- wudfhost.exe
- yourphone.exe
- dwm.exe
- taskhost.exe
- csrss.exe
- system.exe
- regasm.exe
- idle.exe

5 FURTHER DATASET EXPLORATION

Here we delve a little deeper into the obtained VirusTotal dataset. This time, our focus was not on the URLs and campaigns but on the 311M individual submissions within the dataset. That is, the raw 311M records into VirusTotal that contained all the URLs, vendor hits, timestamps, and all important content-hashes. The hope is that this would provide readers with a more contextual information.

5.1 General Findings

Finding 10: Most submissions originated from the United States (US). The US leads the way with a quarter of all submissions (78,687,777 – 25.25%), followed by South Korea (47,973,096 – 15.4%), Japan (38,578,771 – 12.38%), and Germany (32,681,855 – 10.49%). As the world’s technological hub it is perhaps reasonable that most submissions originate from the US from businesses that use VirusTotal via its API. Also, with populations that are more technology literate, more people in such developed countries are aware of, and make use of, the manual submission features of VirusTotal.

Finding 11: Majority of submissions were automated, with a large percentage from a select few. By grouping on the interface sub-field within each submission, we found an overwhelming 91.28% (284.4M) of them came thru the API category, with email, web, community respectively scoring 7.33% (22.8M), 1.06% (3.3M) and 0.32% (1M). This shows that most use of the platform is automated, perhaps from organisations that have commercial agreements with VirusTotal. Further, based on the submitter_id field, we found that just two organisations were responsible for 50.42% of all submissions, with both contributing ~47M each. In fact, the top ten organisations were responsible for close to half (48.34%) of all submissions.

Finding 12: 58.98% of submissions were unflagged. We observed that 183,785,016 submissions were not flagged at all by any of the vendors within VirusTotal. Going by the reasonable premise that only suspicious URLs are being submitted, this means that more than half of submissions were either truly benign or undetected malicious. In the first case, this perhaps signals the need for more work such as [4, 40] to understand why users may deem benign URLs as suspicious. In the second case, this points to the need for more effective methods to better detect attacks as they occur.

Finding 13: 17M unique pieces of content were flagged. By grouping on submissions’ content hashes, we identified that 20.6% of unique hashes (17,116,732) within the dataset were potentially malicious. This was pulled from the corresponding 41.02% of submissions (127,803,111) where at least one vendor had flagged an attached URL as malicious. A majority of the 17M consisted of single submissions, where a unique hash had only one submission attached to it.

5.2 Vendor Effectiveness

Submitted URLs are analysed in parallel by a median of 72 security vendors to determine if they’re benign or malicious. As seen in Table 1, each submission has a positives field that contains the number of vendors that have flagged it. In Figure 5, we delved deeper into the distribution of submissions with respect to vendor flags, and observed that detection performance varies significantly with most submissions being flagged by only a select few vendors.
Finding 14: We observed that around 98.27% (125.6M) of all flagged submissions were detected by 10 or fewer vendors. This indicates that vendor detection performance is highly skewed and only a few of them are effective. However, it must be noted that vendors have different specialisations, with some better designed to detect phishing and others customised to detect other malware. We observed that only seven submissions were detected by the observed maximum of 21 vendors. More specifically, only 1.69% of all flagged submissions were detected by more than 10 vendors (2.1M out of 127.8M), which indicates that a select few were responsible for most detections.

Finding 15: We observed that on average around 4 security vendors detected most of the flagged submissions. This implies that if a URL is flagged by at least four vendors, it is reasonable to conclude that it is malicious. This may be of some assistance to SOC admins when selecting a threshold for labelling URLs based on VT vendors. For example, if at least n number of vendors has flagged a URL, it is likely malicious. If an overly-restrictive threshold (e.g., n=20) is set, it may result in significant number of missed detections. Thus, appropriate thresholds may be set based on Figure 5.

5.3 Time to n Vendor Flags
Within each submission contains a first_seen field. We then asked ourselves: what is the mean time required for submissions to be flagged by n number of vendors? That is, how quickly do vendors catch up in flagging URLs after it was first submitted.

Finding 16: The number of vendors flagging submissions is inversely correlated to the time these submissions were first reported. More specifically, if a large number of vendors detected a particular set of submissions, it is more likely that these submissions are recently reported to the VT. Conversely, if submissions were flagged by only a few, it is more likely that they were reported comparatively long ago. Intuitively, it is expected that more vendors would flag malicious URLs over time but surprisingly, this relationship between the time and vendors is quite opposite. Figure 6 shows time-vendor distribution for the flagged submissions.

Finding 17: Submissions flagged by 12 or more vendors are detected significantly earlier compared to others. There could be two possible reasons for this. First, it indicates that some security vendors are trained on specific types of attacks and consequently, they are able to flag URLs in submissions promptly. We see this in Figure 6 for the time distribution of 12 to 21 vendors. Second, vendors may rely on others when flagging URLs. For example, if 10 vendors flag a particular URL, others may also follow suit.

6 CONCLUSION
We hope that the 17 findings presented can aid SOC personnel rapidly triage and act upon the most pertinent threats. Traditional approaches must be improved as we cannot rely on the wholesale lumping of URLs into binary benign or malicious classifications. For deep-learning based detectors, threat intelligence such as this could help tease out features and connections amongst disparate URLs. The already bewildering array of tricks employed by attackers are ever evolving, as seen in the range of domains, sub-domains and suffixes used to deceive both defences and end-users in Section 4.2. We also note the widely variant URL lengths and propensity for longer URLs as seen in Section 4, and the increased sophistication with fileless malware in Section 4.3.

However, such malicious attacks ultimately hinges on a singular pivot point: the humble URL. This is true regardless of the point wherein the URL was introduced into the SIEM, be it within a suspicious email to an employee, or as an outgoing suspicious request from a possibly compromised corporate device. Accordingly, measuring suspicious URLs as part of coherent campaigns allows for more refined, robust and timely detection protocols within SOCs. Perhaps a key takeaway is that we can already achieve much better threat intelligence by simply inspecting content hashes, without necessarily and immediately leaping to deep-learning or other more advanced heuristic approaches.

Thus in pursuit of the above goal, the work herein is a culmination of extensive research and analysis into 311M records submitted to VirusTotal a 2-month period ending Jan 2020. While there exist...
extensive studies on malicious URLs, there remains a gap in our understanding of them as concerted campaigns. Our work clusters URLs via their metadata and doubly verifies them as malicious before further analysis. This offers practical insights into the nature of campaigns along dimensions such as individual URL attributes and targeted victim brands. We also explored case studies that illustrate various adversarial techniques that can circumvent defences. Future work includes looking at measures to further group campaigns together. That is, to see if there are avenues for us to connect disparate campaigns into a meta-campaign attributable to a common attacker.

ACKNOWLEDGMENTS

The work has been supported by the Cyber Security Research Centre Limited whose activities are partially funded by the Australian Government’s Cooperative Research Centres Programme. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. This research was supported by use of the Nectar Research Cloud and by the Tasmanian Partnership for Advanced Computing (TPAC). The Nectar Research Cloud is a collaborative Australian research platform supported by the NCRIS-funded Australian Research Data Commons (ARDC).

REFERENCES

[1] Neda Alizafar, Yuan Mao, Sandra Michalka, Qing Liu, and Hua Wang. 2020. From logs to Stories: Human-Centred Data Mining for Cyber Threat Intelligence. IEEE Access (8, 2020), 19089–19099. https://doi.org/10.1109/ACCESS.2020.2966760

[2] Mohinuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. 2016. Deep Approaches on Malicious URL Classification. In 20th International Conference on Digital Forensic Research and Law Enforcement (DFRLoG). 1–12. https://doi.org/10.1145/3063167.3063197

[3] Olusola Akinrolabu, Ioannis Agrafiotis, and Arnau Erola. 2018. The challenge of URLs via their metadata and doubly verifies them as malicious beyond understanding of them as concerted campaigns along dimensions such as individual URL attributes and targeted victim brands. Our work clusters URLs via their metadata and doubly verifies them as malicious before further analysis. This offers practical insights into the nature of campaigns along dimensions such as individual URL attributes and targeted victim brands. We also explored case studies that illustrate various adversarial techniques that can circumvent defences. Future work includes looking at measures to further group campaigns together. That is, to see if there are avenues for us to connect disparate campaigns into a meta-campaign attributable to a common attacker.
