#Twiti: Social Listening for Threat Intelligence

Hyejin Shin
Samsung Research
Republic of Korea
hyejin1.shin@samsung.com

WooChul Shim*
Samsung Research
Republic of Korea
woochul.shim@samsung.com

Saebom Kim
Samsung Research
Republic of Korea
sae-bom.kim@samsung.com

Sol Lee
Samsung Research
Republic of Korea
soll.lee@samsung.com

Yong Goo Kang
School of Cybersecurity, Korea University
Republic of Korea
yonggoo@korea.ac.kr

Yong Ho Hwang
Samsung Research
Republic of Korea
yongh.hwang@samsung.com

ABSTRACT

Twitter is a popular public source for threat hunting. Many security vendors and security professionals use Twitter in practice for collecting Indicators of Compromise (IOCs). However, little is known about IOCs on Twitter. Their important characteristics such as earliness, uniqueness, and accuracy have never been investigated. Moreover, how to extract IOCs from Twitter with high accuracy is not obvious. In this paper, we present *Twiti*, a system that automatically extracts various forms of malware IOCs from Twitter. Based on the collected IOCs, we conduct the first empirical assessment and thorough analysis of malware IOCs on Twitter. Twiti extracts IOCs from tweets identified as having malware IOC information by leveraging natural language processing and machine learning techniques. With extensive evaluation, we demonstrate that not only can Twiti extract malware IOCs accurately, but also the extracted IOCs are unique and early. By analyzing IOCs in Twiti from various aspects, we find that Twitter captures ongoing malware threats such as Emotet variants and malware distribution sites better than other public threat intelligence (TI) feeds. We also find that only a tiny fraction of IOCs on Twitter come from commercial vendor accounts and individual Twitter users are the main contributors of the early detected or exclusive IOCs, which indicates that Twitter can provide many valuable IOCs uncovered in commercial domain

CCS CONCEPTS

• **Information systems** → **Information extraction**; • **Human-centered computing** → **Social networking sites**.

KEYWORDS

IOC, Twitter, open source threat intelligence, threat hunting

*Corresponding Author

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3449797

ACM Reference Format:
Hyejin Shin, WooChul Shim, Saebom Kim, Sol Lee, Yong Goo Kang, and Yong Ho Hwang. 2021. #Twiti: Social Listening for Threat Intelligence. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3442381.3449797

1 INTRODUCTION

Malware attacks are increasing every year [1, 7, 14]. In particular, malware distribution through websites is rapidly increasing [33]. As seen in Dyn attack [41] and Garmin ransomware attack [32], malware can quickly spread out and its damage can be catastrophic. Considering its risks, prevention is the best defense. Indicators of Compromise (IOCs) are by far the key to defending against malware, although there are some prediction based malware detection solutions. IOCs are forensic artifacts of a cyber attack, so they enable detecting intrusion attempts or any other malicious activities on a system or network. When up-to-date IOCs are timely fed, they play a critical role in protecting the system or network from future attacks. Examples of IOCs include MD5 hashes of malicious files, IP addresses, URLs or domains of botnets, and file names.

Most organizations subscribe threat intelligence (TI) feeds to receive malware IOCs, but a single feed is not enough. Many antivirus solutions and commercial TI feeds often do not reflect IOCs for new or ongoing attacks immediately [42, 43, 47]. Furthermore, regardless of public or commercial TI feeds, the coverage of one feed is incomplete [36, 42, 47]. For these reasons, many industries and security professionals enrich IOCs through open source threat intelligence [29, 36]. According to a 2019 survey among 1,908 IT and security practitioners in North America and the U.K. [29], at least 37% of respondents say that their organizations use public TI feeds together with commercial feeds (41% of respondents say that their organizations use one paid TI feed while 78% respond to using more than one TI feed).

There are lots of public sources to gather malware IOCs. The most easily accessible sources are public malware blocklists such as Feodo tracker [2] and AlienVault IP reputation [5]. Security vendor blogs are another common sources for IOC mining [21, 22, 43]. Security mailing lists, security forums, and dark web are also often used for IOC hunting [3, 21, 39, 40]. Among many public sources, Twitter guarantees volume, timeliness, and diversity in attacks [49]. It brings vast quantities of contents from across the web by linking tweets to external sites, which enables Twitter to

cover plenty of fresh IOCs coming from various sources such as security vendor blogs, honeypots, and malware sandboxes. This makes many security vendors utilize Twitter in practice for IOC hunting [11, 21, 22, 28].

However, mining IOCs with high accuracy from Twitter is not obvious due to the unique characteristics of Twitter such as short text, non-standard language, and diverse external sources linked to tweets. There are some open systems [13, 23] that gather IOCs from Twitter. But, as we show later in Section 4.3, our experiments show that the coverage and accuracy of both systems are unsatisfactory to make use of IOCs on Twitter. Hence, we develop Twiti, an automatic IOC extraction system for Twitter. Twiti identifies tweets that are likely to contain malware IOCs using a tweet classifier and a selected list of external sources. It then extracts IOCs both from tweets and external links in tweets. This approach enables Twiti to collect a large number of IOCs with high accuracy.

Moreover, despite the popularity of Twitter as a data source for IOCs, little is known about IOCs collected from it – how many IOCs are on Twitter, how fresh they are, how accurate they are, how unique they are compared to other public or commercial TI feeds, what kinds of malware IOCs are reported, who reports exclusive IOCs, what fraction of IOCs on Twitter can be used for any purpose, how many IOCs can be obtained from external links, and so on. To answer these questions, we collect malicious file hashes as well as IP addresses, domains, and URLs that are related to malware through Twiti. We then evaluate volume, latency, accuracy, and exclusiveness. We finally analyze the characteristics of the collected IOCs in various aspects from data source, file type to malware type in order to provide insight about IOCs on Twitter.

Our contributions are summarized as follows:

- We develop a high-performance IOC extraction system that mines up-to-date malware IOCs from Twitter. The comparisons with the largest public TI feed (AlienValut OTX Pulse), some public IP blocklists, and the largest threat information aggregator (VirusTotal) show that Twiti achieves **high accuracy** (a precision of 0.93 for file hashes and a precision of 0.92 for URLs), **high uniqueness** (62.74% exclusiveness for file hashes with respect to AlienVault OTX Pulse, 34.45% exclusiveness for URLs with respect to VirusTotal, and more than 90% exclusiveness for IP addresses with respect to popular IP blocklists), and **low latency** (59.87% of file hashes detected 3.5 days earlier than AlienVault OTX Pulse and 51.81% of URLs detected 1.7 days earlier than VirusTotal) (§ 4).
- We demonstrate that **Twitter is a proper TI source for defending against ongoing malware attacks.** Against Emotet that has tremendous variants and persistently appears, Twiti collected Emotet hashes in volume (approximately 16,000 hashes over 3 months) very accurately (a precision of 0.95), exclusively (77.06% of Emotet malware in Twiti were unique compared to AlienValut OTX Pulse), and early (92.09% of Emotet hashes in Twiti detected 1.8 days earlier than AlienVault OTX Pulse) (§ 4.2.1). For malware-related URLs that well reflect ongoing malware attacks due to their transient nature, Twiti detected them almost two days ahead of several website scanning engines like Google Safe Browsing. Also, nearly one third of the malware-related URLs detected by Twiti were not found within 30 days by website scanning engines (§ 4.2.2).
- We believe **we are the first group to thoroughly examine the characteristics of IOCs on Twitter.** We provide information on which source on Twitter reports IOCs early, what fraction of IOCs collected from Twitter can be used without data use restriction, and who reports IOCs exclusively and early (§ 5.1). We also analyze what kinds of malware hashes or which types of malicious URLs can be obtained from Twitter (§ 5.2).

2 TWITI: DESIGN AND IMPLEMENTATION

Figure 1 illustrates the architecture of Twiti. Twiti consists of three steps – data collection, relevant tweet selection, and IOC extraction. Twiti aims to collect malware-related IOCs as many as possible with high accuracy. To achieve it, we carefully designed our data collector and IOC extractor by performing a pilot study in November 2019, as presented in Section 3. The source code of Twiti is available at `https://github.com/Samsung/Twiti`.

2.1 Tweet Collector

To maximize the number of IOCs to be collected, Twiti collects data primarily by keyword tracking using Twitter search API [24] and secondarily by user tracking using timeline API [25]. We have tracked 35 keywords that are likely to appear with malware IOCs. Examples of keywords are "malware", "ransomware", "botnet", "spyware", "adware", "malspam", "iocs", and "virustotal.com". In addition, we have tracked 146 Twitter users who consist of 86% security experts, 12% security vendors, and 2% other security organizations. Note that among 125 security experts, 67% introduced themselves as malware analyst, malware researcher, threat hunter, or threat intelligence researcher in their profiles. Also, note that Twiti collects the original tweets of retweets and extracts IOCs from them.

2.2 Relevant Tweet Selector

A simple extraction of IOCs with pattern matching causes many false positives. Most tweets include links for their own tweets or references (e.g., https://t.co/qQdme1Buxh). Several tweets mentioning software versions match to IP pattern (e.g., Tuleap 9.17.99.189). Some tweets mention hashes for referring commit IDs or blockchain transactions. To reduce such false positives, Twiti first handles links in tweets and then classifies tweets to filter out those without IOCs.

2.2.1 Tweet preprocessor.

Shortened URL remover. All links (URLs) posted in tweets are automatically shortened by Twitter's t.co service. Since links converted by Twitter are checked against potentially dangerous sites, we delete "http://t.co" links from texts in order to avoid falsely detecting benign URLs in tweets as IOCs. Albeit this process, some links shortened by other URL shorteners sometimes remain in tweets. Hence, we additionally delete short URLs whose domains are "bit.ly", "tinyurl.com", "buff.ly", "goo.gl", "youtu.be", or "ow.ly".
Regex checker. After shortened URLs are removed, we check if there are any terms matching with regular expressions for hashes, IP addresses, domains, and URLs in each tweet.

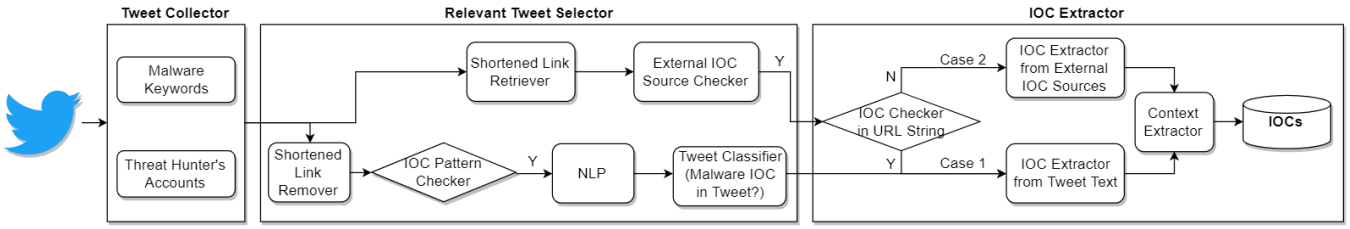

Figure 1: The system architecture of Twiti for an automatic IOC mining from Twitter.

Text preprocessor. For each tweet passing through Regex checker, the following natural language processing (NLP) is applied to extract features for a classifier:

(1) All the types of hashes are replaced with "[hash]". Terms for IP addresses, URLs, domains, file names, file paths, and emails are also replaced by "[ip]", "[url]", "[domain]", "[filename]", "[filepath]", and "[email]", respectively. Note that all the defanged URLs, IP addresses, and domains are transformed into their representative tokens like "[url]". Twitter handles and CVE IDs are replaced by "[username]" and "[cve]" as well. All the numeric numbers are replaced by "[num]".
(2) Named entity recognition (NER) is applied to each tweet. The words tagged as malware are replaced with "[malware_name]".
(3) Twitter handles in pre-texts and post-texts are removed.
(4) Unicode characters and symbols not used in IOCs are removed.
(5) The tweet is lowercased. The tracked keywords and their aliases are replaced by single representative terms in the form of a single token. For example, "c&c", "cnc", and "command-and-control" are replaced by "c2".
(6) The tweet is tokenized and lemmatization is applied to each word in order to represent the inflected forms of a word as a single word. Stopwords are removed. The words consisting of a single character, "[username]", and "[num]" are deleted.

Note that the existing NER tools like NLTK [44], CoreNLP [45], and twitter_nlp [48] are not trained in cybersecurity domain. So, we trained the Bert model [37] with tweets mentioning cybersecurity events and used it in the step (2). The details of our Bert-based NER can be found at `https://github.com/Samsung/Twiti`.

2.2.2 *Tweet classifier.*

We developed a high-performance tweet classifier that determines whether a tweet contains IOCs or not. We hereafter call tweets either IOC tweets or non-IOC tweets according to whether they include IOCs or not.

Dataset. To build an IOC tweet classifier, we had collected tweets containing IOC patterns from January to September 2019. We could collect 21,937 tweets during the period. After removing similar tweets whose Jaccard similarity is greater than 0.70, 5,675 tweets were remained. Three security experts manually annotated whether each tweet includes any IOCs or not. There were 3,007 IOC tweets and 2,668 non-IOC tweets.

Features. We considered the followings as initial features:

- *Defanged IOCs*: This feature checks if there is at least one defanged IOC in each tweet. When posting IOCs on Twitter, the defang techniques are commonly applied to IP addresses, URLs, and domains in order to prevent accidental exposure to malicious active content. Examples of such tweets are *"#gandcrab @ hxxp://92.63.197.106/c.exe"*, *"#RoamingMantis new landing pages: 67[.] 198.129.27 ..."*, *"#darkcomet /elumadns.eluma101.com ..."*, *"This app impersonate ... #c2 hold[.]jcgloball[.]org:11880"*.
- *Contextual n-grams*: These are contextual words surrounding pivotal words for IOCs. The pivotal words we used are the tracked keywords (e.g., "malware", "ransomware", "botnet"), "[hash]", "[ip]", "[url]", "[domain]", and "[malware_name]". It is intuitively clear that the words around patterns of interest would be very different in between IOC and non-IOC tweets. For example, "version [ip]", "up to [ip]", "before [ip]", "prior to [ip]", and "commit [hash]" clearly show up in tweets about software vulnerability while "hash [hash]", "c2 [url]", "c2 [ip]", "botnet c2", "from [ip]", "ransomware [hash]", "[filename] [hash]", and "[malware_name] md5s [hash]" definitely belong to IOC tweets. To extract such contextual features, we first apply the text preprocessing (1)-(5) to each tweet. We then extract bigrams and trigrams consisting of a target word and 1-2 words to its left and right sides.
- *Bag of words*: Words co-occurred with IOCs would also differ from those in non-IOC tweets. For example, "c2", "md5s", "yara", "botnet", "[malware_name]", "ransomware" obviously appear more in IOC tweets. On the contrary, "[cve]", "csrf", "0daytoday", "vulnerability", "xss", and "sql" are less likely observed in IOC tweets. the text preprocessing (1)-(6) are applied to extract words. We then delete common English words. By considering lemmatized words as features, we can consider word variation which cannot be taken into account in contextual features.

Note that all the features here are binary features. That is, each feature takes the value 1 if it is in a tweet, and 0 otherwise.

Feature selection. Not all the features are important for classification. We chose features that differentiate IOC tweets from non-IOC tweets using the mutual information (MI). For a feature X and the class label $Y \in \{\text{IOC tweet, non-IOC tweet}\}$, the mutual information of X and Y is computed by

$$\mathrm{MI}(X,Y) = \sum_{x\in\{0,1\}} \sum_{y\in\{0,1\}} P_{X,Y}(x,y) \log\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right),$$

where $P_{X,Y}$ is the joint distribution of X and Y, and P_X, P_Y are the marginal distributions of X and Y respectively. MI measures how much knowing X reduces uncertainty about Y and vice versa. For example, if X and Y are independent, then knowing X does not give any information about Y, so their MI is zero. Thus, MI enables us to pick the features that are helpful to distinguish IOC tweets from non-IOC tweets. We take the words and n-grams whose MI is larger than 0.0002. The threshold was chosen to maximize the prediction performance of a classifier.

Classifier. There were 22,316 initial features. After feature selection, 1,456 features were retained. They contain 483 words (unigrams) and 972 bigrams and trigrams. We considered 3 classifiers – logistic regression, random forest, and XGBoost. We evaluated these classifiers with the dataset consisting of 3,007 IOC tweets and 2,668 non-IOC tweets using a 5-fold cross validation. We chose the random forest classifier since it showed the best performance – a precision of 0.95 and a recall of 0.96. We present ROC curve of 3 classifiers in Figure 2 and examples of important features of the random forest classifier in Table 1.

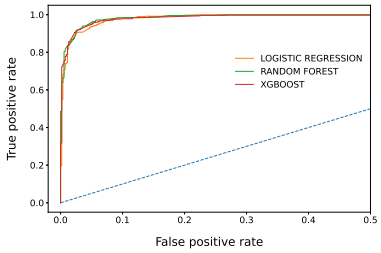

Figure 2: ROC curves of logistic regression, random forest, and XGBoost classifiers. Their AUC values are 0.988, 0.989, and 0.988.

Table 1: Important features of the random forest classifier

Features	Importance*	# tweets	Features	Importance	# tweets
is_defanged	100	1682	payload	6.99	179
[malware_name]	87.32	1182	phish	6.18	57
[hash]	83.75	1390	version [ip]	6.17	169
[cve]	57.09	1316	[url] [url]	6.16	197
[ip]	56.51	3177	to [ip]	5.89	256
[url]	49.14	1473	malspam	5.62	176
c2	27.68	492	[filename]	5.18	431
malware	19.21	424	commit [hash]	4.89	52
[ip] and	13.69	400	0daytoday	4.77	101
botnet	12.79	143	attacker	4.75	172
[domain]	12.43	333	[hash] [hash]	4.42	117
phishing	11.95	201	hash [hash]	4.30	147
from [ip]	10.84	170	bgp	4.15	102
ip [ip]	9.47	118	md5 [hash]	4.02	153
[filename] [hash]	8.26	167	buffer	3.83	96
ip	7.87	280	up to [ip]	3.53	100
md5	7.85	210	[ip] allows	3.37	135
before [ip]	7.72	219	c2 [ip]	3.16	77
opendir	7.50	176	c2 [url]	2.97	139
ransomware	7.41	153	[hash] on	2.85	119

*Gini importance score that is normalized so that the largest score is 100.

2.2.3 *External link checker.*

Due to text brevity of tweets (280 character limit), users often share detailed information via external links. Hence, we built a list of the external sources that provide a large number of IOCs with small false positives by analyzing the external links in tweets from the pilot study. The selected external sources are presented in Table 9. Since all links in tweets are shortened by Twitter, Twiti retrieves the full URLs of "http://t.co" links from Twitter API. It then checks if the full URLs are from the selected external sources.

2.3 IOC Extractor

On Twitter, there is a variety of threat-related information ranging from vulnerabilities, exploits, and malware to anomalous cyber activities [49]. However, the level of detail of such information varies from author to author. Some Twitter users post C&C servers or other valuable IOC information like IP addresses, URLs, and file hashes. On the other hand, other users share their findings or experiences without much detail. According to the level of information detail, approaches to hunting IOCs from Twitter differ. In Twiti, IOC extractor runs into the following 2 cases:

- Case 1: IOCs in tweets.
- Case 2: No IOCs in tweets, but IOCs in external links.

IOC extraction from tweets. Twiti first looks for IOCs in tweet texts by pattern matching with regular expressions. However, certain types of IOCs such as URLs and IP addresses are often defanged to avoid inadvertent clicking on a malicious link. From our evaluation, we found that 38% of the collected IPs were defanged and 73% of the collected URLs were defanged. This shows that there are more challenges on Twitter in handling defang techniques than in security blogs, forums, and mailing lists. Twiti detects defanged IOCs by using various defang techniques in open-source IOC extractors [15, 16] together with more patterns for defanged URLs we added to extend the detection coverage. Twiti also collects file hashes, IP addresses, and domains from link texts themselves. Recall that Twiti deletes "http://t.co" links from texts before pattern matching, though they are parts of tweets. However, from our external link analysis, we observed that some types of IOCs are embedded in the given link texts of malware analysis services. For example, *"https://www.virustotal.com/gui/ip-address/78.155.199.119/detection"*. Thus, Twiti directly extracts those IOCs from the given links.

IOC extraction from external sources. Twiti collects IOCs from the external sources when the links in tweets are in the selected list given in Table 9. To choose external sources providing a large number of IOCs with small false positives, we analyzed the links embedded in tweets collected during November 2019. From our analysis, we found that security vendor blogs, malware analysis services, and Pastebin.com are major sources of IOCs. We separately develop IOC extractors for different types of data sources as follows:

- Pastebin.com: We observed that Pastebin.com is one of top external links given in tweets. It is a website where users can store text online. As we show later in Section 4, many IOCs Twiti collects were from it. In Pastebin, there are various types of information from source code snippets, leaked credentials to IOCs. Thus, for IOC collection, searching all the links for Pastebin.com in tweets is not a good idea. Hence, we analyzed the words co-occurring with Pastebin.com and extracted the top 50 words after applying the text preprocessing (1)-(6). After manual review, we finally chose 18 words.

Examples of such words are "malware", "ransomware", "trojan", "botnet", "[malware_name]", "c2", "ioc", and "payload". Twiti gathers IOCs from Pastebin when those words show up together with the Pastebin.com links.
- Malware analysis services: We observed that IOCs in tweets are often given together with the links of analysis reports. From our external link analysis, we observed that 57% of analysis reports posted in tweets were from VirusTotal, 33% from Any.Run [6], 7% from urlscan.io [27], and 3% from the remaining malware analysis services. Many of them contain IOCs in the given link texts, but some provide IOCs in their sites. In the latter case, Twiti collects IOCs using their APIs. Note that, although we observed that many malicious file hashes earlier than VirusTotal are often reported via app.any.run, Twiti cannot collect IOCs from Any.Run since there is no public API offered by it.
- Security vendor blogs: We observed more than 100 security vendor blogs from our external link analysis. Each vendor has its own format in providing IOCs. Thus, dedicated parsers need to be developed for each blog.
- Other than those mentioned above, Twiti collects IOCs from AlienVault OTX [20] using API.

Remark that almost all security vendor blogs strictly restrict use of their data in their Terms of Service. So, Twiti collects the data from major 10 security vendor blogs in the number of IOCs among hundreds of vendor blogs for informational use only to give insights on IOC data collected from security vendors. Since the purpose of our work is not a performance evaluation between security vendor blogs, we anonymize their names in Table 9.

3 DESIGN CHOICE

The followings are our design choices for Twiti to collect malware IOCs as many as possible with small false positives.

Data collection method. There are two ways to collect data from Twitter – (i) keyword tracking and (ii) user tracking. To determine a data collection method for Twiti, we experimented how different the number of IOCs is in between two methods. For the experiment, we tracked 35 keywords and 82 Twitter users during November 2019. We observed that 36.2% of the collected IOCs were from keyword tracking, 25.6% from user tracking, and 38.2% from both. Thus, we decided to leverage both ways to maximize IOC collection. Since keyword tracking pulls IOCs much more and is easier to expand, Twiti uses keyword tracking as a primary data collection method and user tracking as an auxiliary method.

Selection of keywords. We selected the keywords that are likely to co-occur with IOCs, but do not make too much noise. We extracted the top 100 words that appear more in IOC tweets than non-IOC tweets using the dataset in Section 2.2.2. We applied the text preprocessing (1)-(6) in Section 2.2.1. We then removed the common words in Twitter and the normalized words like "[malware_name]" and "[cve]". After deleting general words that could cause many false positives, we obtained 35 words.

Selection of Twitter users. In order for the user-based data collection to be complementary to the keyword-based data collection, we chose Twitter users who met any of the following conditions:

(1) Is a user often mentioning IOCs without the keywords above?
(2) Is a user the original tweet author of retweet containing IOCs or in the threads of discussion about IOCs?
(3) Is a user a contributor of IOCs?
(4) Does a user's profile include the words like "malware", "ransomware", "threat hunter", and "threatintel"?

We collected such users by analyzing the dataset in Section 2.2.2 and their profiles. We extracted the authors who had created at least one IOC tweet without the keywords and whose accounts were active. In addition, we extracted users in pre-texts and post-texts of IOC tweets since we observed that users located at the beginning and end of IOC tweets fall into the conditions (2)-(3). We then retained the users whose appearances are statistically significantly larger in IOC tweets than non-IOC tweets. Finally, we analyzed the account profiles of the collected users and found that many of them introduced themselves as malware analyst, malware researcher, threat hunter, or threat intelligence researcher. We extracted a few important words from their profiles and then collected more Twitter users including those words. After all the above processes and manual review, 146 Twitter users were selected.

Selection of external sources. We analyzed the links embedded in IOC tweets collected during November 2019. We obtained 25,437 unique reference URLs consisting of 5,605 unique domains. Among them, we chose the top sites for IOC collection. Note that, among 25,437 external links, 6.2% were from malware analysis services, 4.2% from security vendor blogs, 1.4% from Pastebin.com, and 0.15% were pulses in AlienVault OTX.

4 EVALUATION

4.1 Evaluation Setup

Evaluation metrics. To evaluate the performance of Twiti, we measured volume, exclusiveness, latency, and accuracy by comparing IOCs collected by Twiti to the selected reference sources. For each type (e.g., file hash) of indicators, we defined

- **Volume** as the total number of indicators in a feed during the evaluation period.
- **Exclusiveness** as the proportion of indicators in Twiti that were not in a reference source during their lifetime. It is formally given as $|\text{Twiti} \backslash A| / |\text{Twiti}|$ for a reference A.
- **Latency** as the elapsed time between the first detection of an indicator by Twiti and its first appearance in a reference source within its lifetime.
- **Accuracy** as the proportion of indicators in a feed that were truly malicious, which corresponds to precision.

Coverage, the proportion of the intended indicators that are captured by a feed, is an important performance metric. However, it is difficult to measure coverage in the absence of ground truth on all ongoing threats [36]. So, as in Bouwman et al. [36], we instead measured the proportion of indicators in a feed that were captured by Twiti when the entire set of indicators in the feed was available.

Reference sources. Table 2 summarizes the reference sources we used for evaluation. We used VirusTotal as a ground truth to measure accuracy of hashes and URLs. We also used VirusTotal to measure exclusiveness and latency for all IOC types. VirusTotal is not only a service that analyzes suspicious files and URLs to detect malware, but also it is the largest TI feed that is supported by 72 antivirus engines and 68 website/domain scanning engines and

blocklists [30]. Thus, high exclusiveness and low latency compared to VirusTotal would be a good indicator of the strength of Twiti as a TI feed. Note that we used the VirusTotal private API v3.0 to get reports on file hashes, URLs, IP addresses, and domains for research purposes. We additionally used the following references. (i) For file hashes, we compared Twiti with AlienVault OTX Pulse and MalwareBazaar [18]. Both of them are not contributors to VirusTotal. AlienVault OTX is the largest open threat exchange platform where anyone can subscribe IOCs via pulse subscription. MalwareBazaar claims that two-thirds of their samples are not detected by VirusTotal. (ii) For domains, we used top 25k domains in Alexa top 1M [4], Cisco Umbrella top 1M [9], and Majestic 1M [17] data to check how many benign domains were reported as malicious. For each 25k domain set, we used the domains that have been continuously appeared over the evaluation period since there could be malicious domains in the list for a while. (iii) For IP addresses, we compared Twiti with some public IP blocklists that are related to malware. The selected public IP blocklists were AlienVault IP Reputation, Bambenek_c2, Feodo Tracker, SSL Blacklist, and a Mirai-related feed. To measure accuracy, we built an allowlist of IP addresses with the above-mentioned top 25k domain data and major content delivery network (CDN) services (AWS CloudFront, CloudFlare, Fastly, EdgeCast, and MaxCDN). Since VirusTotal includes almost all of popular URL and domain blocklists that are open to public, we compared URLs and domains in Twiti only with VirusTotal.

Table 2: Reference sources for evaluation

IOC Type	Exclusiveness	Latency	Accuracy
File hash	AlienVault OTX Pulse, MalwareBazzar, VirusTotal		VirusTotal
URL	VirusTotal		VirusTotal real-time scan
Domain	VirusTotal		Popular domains
IP	IP Blocklists, VirusTotal		Popular domains, CDN

Dataset and IOCs for evaluation. We ran Twiti on a daily basis over 978,414 tweets that we had collected by tracking 35 keywords and 146 users from February to April 2020. After deleting duplication and filtering through regex, tweet classifier, and external link checker, 17,904 tweets classified as IOC tweets and 9,372 tweets including the external links in our watchlist remained. From those tweets, Twiti collected 32,200 unique file hashes, 18,718 unique URLs, 70,515 unique IP addresses, and 11,060 unique domains. We evaluated all the file hashes we had collected for 3 months. Meanwhile, we evaluated URLs, IPs, and domains only for the month of April because a daily tracing of a large number of URLs, IPs, and domains easily exceeded the daily query limit of VirusTotal API. For the same reason, we compared Twiti to AlienVault OTX Pulse only for file hashes.

4.2 Evaluation Results

4.2.1 File hashes.

On each day, Twiti collected file hashes that have not seen before. Table 4 shows the evaluation results of file hashes collected by Twiti for 3 months.

Volume. Twiti collected 32,200 file hashes over 3 months, where 20,837 hashes were collected in February, 5,306 hashes in March, and 6,057 hashes in April. They consisted of 10,022 MD5 hashes (31.1%), 2,024 SHA1 hashes (6.3%), and 20,154 SHA256 hashes (62.6%). By querying them to VirusTotal, we found that 30,207 hashes in Twiti existed in VirusTotal and they correspond to 22,824 unique files. Among them, there were 982 hashes for Android apps, 320 hashes for ELF files, and 33 hashes for iOS apps, which correspond to 712, 227, and 31 files, respectively. Figure 3 shows the daily number of file hashes collected by Twiti. Twiti could stably collect ample IOCs for 3 months, except when a bunch of file hashes came through Pastebin.com. Note that in the first few days of February, 2-3 users shared hundreds to thousands of IOCs via Pastebin.com links. Except those days, 421 file hashes were mentioned daily on average and Twiti could collect 200 new file hashes daily on average during the evaluation period.

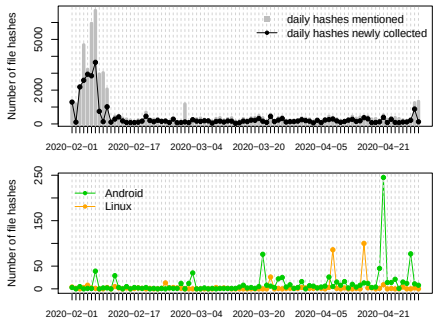

Figure 3: The daily number of file hashes collected by Twiti (Top) and the daily number of Android and Linux file hashes (Bottom).

Exclusiveness. We compared all collected hashes to VirusTotal and AlienVault OTX Pulse using their APIs. We queried hashes to each source and then checked whether they are found in each source or not. We viewed a hash as being in VirusTotal when it was detected as malicious by at least one of 72 antivirus engines. In other words, hashes not in VirusTotal were those either undetected by any engines or not found in VirusTotal. By doing this, we observed that among 32,200 file hashes in Twiti, 7.20% were not in VirusTotal and 62.74% were not in AlienVault OTX Pulse, as of May 1st.

Latency. We defined the first detection time of a file hash by Twiti as its first appearance time in tweets we had collected since February. This meant that all of the file hashes collected on February 1st take their first detection date as February 1st, although they may have appeared earlier on Twitter. Comparing the latency of such file hashes to the reference could mischaracterize the performance of Twiti. Thus, we computed the latency of Twiti only for the file hashes whose first detection date in the reference source was February 1st or after that date. There were 21,175 file hashes in Twiti available for latency comparison with VirusTotal. Among them, 814 file hashes (3.84%) were detected by Twiti 1.2 days earlier than VirusTotal on average (maximum 27.5 days) and 14,052 file hashes (66.36%) were detected within 24 hours from the first detection time of VirusTotal. For comparison with AlienVault OTX Pulse, 8,508 file hashes in Twiti were available. Among them, 5,094 file hashes (59.87%) appeared in Twiti 3.5 days earlier than in AlienVault OTX Pulse on average (maximum 86.2 days). Figure 4 shows the latency distribution of Twiti compared to VirusTotal and OTX Pulse.

Accuracy. Since VirusTotal can have false positives as well as its detection can be late, we queried all the collected hashes at the end of May again. We then measured the proportions of hashes that are labeled as malicious by at least one antivirus engine and trusted software. After doing all of these, 92.86% of file hashes in Twiti were malicious, 0.03% were benign, and 7.11% remained unknown in VirusTotal by the end of May. Among unknown hashes, 10.5% were reported by security vendors, 6.6% were from analysis reports of malware analysis services like Hybrid Analysis [12] and URLhaus [26], 5.4% were reported from tweets with app.any.run [6] results, and 1.9% were reported by honeypot accounts. This means that they were suspicious enough, though they were undetected by any engines in VirusTotal.

Emotet hashes. Emotet malware was discovered in 2014, and it has recently evolved into a threat distributor acting as Malware-as-a-Service by distributing and dropping other banking Trojans such as Trickbot, Ursnif, and Ryuk payload. To defend against massive volume of its variants efficiently, it is important for a TI feed to collect Emotet hashes in volume as early as possible. Twiti can collect malware hashes for Emotet in volume. It collected 16,539 file hashes (corresponding to 11,761 malware samples) co-occurring with the word "emotet" for 3 months. By querying them to VirusTotal, we observed that 95.04% were malicious, 4.95% were remained unknown, and only 1 hash was benign. Twiti showed a higher accuracy for Emotet hashes than other malware hashes. Also, Twiti collected 92.09% of Emotet hashes 1.8 days earlier than AlienVault OTX Pulse and all Emotet hashes 33.3 days earlier than MalwareBazaar. We also measured the overlap of Emotet malware samples among Twiti, AlienValut OTX Pulse, and MalwareBazaar. The result is given in Table 3. Compared to AlienVault OTX Pulse and MalwareBazaar, not only could Twiti collect the largest number of Emotet malware samples highly exclusively (77.06% and 99.09%), but also it could cover one third of malware samples in other public TI feeds.

Table 3: Comparison of Emotet malware samples among Twiti and other public TI feeds

		Volume	Exclusiveness	Overlap with Twiti[1]
Twiti		11,761		
Public TI Feed	AlienVault OTX Pulse	8,125	77.06%	33.20%
	MalwareBazaar[2]	317	99.09%	33.75%

[1] $|A \cap \text{Twiti}|/|A|$ for a feed A. [2]MalwareBazaar started a service from February 13th 2020.

4.2.2 URLs.

The evaluation of URLs is more complicated than file hashes. The owner or content of a URL is varying over time, so it could be malicious one day, but benign the other day. Based on the earlier studies [34, 46], we considered 30 days as the lifetime of malicious URLs related to malware such as malware distribution sites or C&C URLs. Table 5 shows the evaluation results of URLs collected by Twiti for a month using a 30-day window.

Volume. Twiti collected 6,873 malicious URLs during April. The average daily number of URLs was 229. Note that Twiti collected 7,630 URLs in February and 4,911 URLs in March.

Exclusiveness. We compared the collected URLs to VirusTotal. We queried each URL to VirusTotal on a daily basis and checked whether it is malicious or not. To judge whether a URL is malicious, we used the latest scan results of VirusTotal. If the latest scan result (last analysis result) of a URL in VirusTotal was malicious and its scan date (last analysis date) was within last 30 days, then the URL was determined to be malicious. If the latest scan result of a URL in VirusTotal was malicious, but its scan date was before last 30 days, we requested to analyze the URL and decide that the URL is malicious in VirusTotal when the re-scan result was malicious. Otherwise, we determined that the URL was not in VirusTotal. Twiti detected 2,368 URLs not in VirusTotal, which was 34.45% of the collected URLs. We think that the time gap between scan update interval and a relatively short lifetime of malicious URLs makes website scanners fail to detect short-lived malicious URLs, which results in a high exclusive rate of URLs. This high exclusiveness tells that even the largest commercial feed is incomplete, so the aggregation of URLs from multiple feeds is beneficial for preventing the spread of malware.

Latency. The latency of a malicious URL was calculated by the difference between its first detection date in Twiti and its latest scan date in VirusTotal that is valid within last 30 days. Similarly to file hashes, we measured the latency of Twiti for the URLs whose latest scan date in VirusTotal was April 1st or after that date. 4,229 URLs in Twiti were available for latency comparison. Twiti found 2,191 URLs (51.81%) 1.7 days earlier than VirusTotal on average, 1,741 URLs (41.17%) on the same day, and 297 URLs (7.02%) later.

Accuracy. We checked if the collected URLs were truly malicious by making requests for analysis to VirusTotal. However, this analysis request modified the latest scan date, so the result of our latency computation above is distorted. Thus, we performed an additional experiment. From May 1st to 14th 2020, we requested VirusTotal to scan the collected URLs immediately after they were detected by Twiti and then measured what fraction of them were malicious or suspicious in the scan results. During the period, Twiti collected 2,386 URLs. Among them, 1,992 URLs were malicious in the VirustTotal scan results, 72 URLs were suspicious, 317 URLs were clean, and 5 URL sites were not found. Because the website scanners in VirusTotal could not always provide up-to-date results, we queried the clean URLs again at the end of May and observed that 142 clean URLs turned to be malicious 2 weeks later. Thus, 89.44% of 2,386 URLs detected by Twiti from May 1st to 14th were truly malicious. Including suspicious URLs, the overall accuracy of Twiti was 92.45%. Despite a high real-time scan accuracy, Twiti collected 7.33% clean URLs, which makes difficult to use Twiti as an automatic feed. As the real-time web scanners in VirusTotal could produce false negatives, we did false positive (FP) analysis of 175 URLs that were identified as clean by VirusTotal. The FP analysis results can be found at our GitHub repository. We found that (i) Twiti's actual false positives were 98 URLs, i.e., a precision of 95.89%, and (ii) 50% of 98 clean URLs were from Pastebin.com when users post IOCs with reference links. Thus, an allowlist consisting of trustable domains in cybersecurity (e.g., virustotal.com, app.any.run, urlhaus, abuse.ch) could eventually improve Twiti's precision to 97.53%.

4.2.3 IP addresses.

IP addresses have the time-varying property like URLs. Many recent studies [42, 47] assume the lifetime of malicious IPs to be 30 days. We also use a 30-day window for our evaluation. Table 6 shows the evaluation results of IP addresses collected by Twiti for a month.

Table 4: Evaluation results of file hashes collected by Twiti

Volume (3 months)	Exclusiveness		Mean Latency (days)				Accuracy		
	AlienVault OTX Pulse	VirusTotal	AlienVault OTX Pulse		VirusTotal		True positives[1]	Unknown[2]	False positives[3]
			Twiti earlier	Twiti later	Twiti earlier	Twiti later			
32,200	62.74%	7.20%	3.5 (59.87%)[4]	2.1 (40.13%)	1.2 (3.84%)	3.1 (96.15%)	29,902 (92.86%)	2,288 (7.11%)	10 (0.03%)

[1] Detected as malicious in VirusTotal, as of May 31st. [2]Not found or undetected by any antivirus engines in VirusTotal, as of May 31st.
[3]Trusted software in VirusTotal, as of May 31st. [4]The fraction of file hashes in Twiti that were detected earlier than AlienVault OTX Pulse.

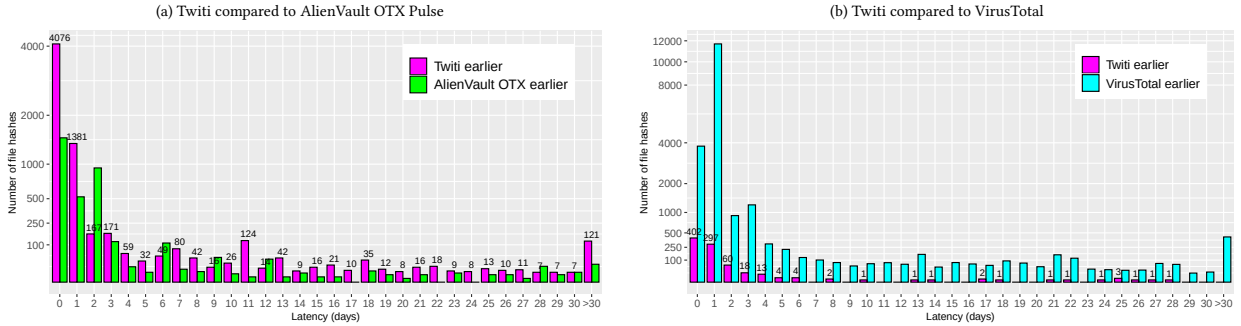

Figure 4: Latency distribution of file hashes collected by Twiti compared to (a) AlienVault OTX Pulse and (b) VirusTotal.

Table 5: Evaluation results of URLs collected by Twiti

Volume (one month)	Exclusiveness	Mean Latency (days)		
		Twiti earlier	Same day	Twiti later
6,873	34.45%	1.7 (51.81%)[1]	0 (41.17%)	5.8 (7.02%)

Accuracy (VirusTotal scan results[2])				
Malicious			Suspicious	Clean
Malware	Phishing	Others		
67.69%	18.31%	3.44%	3.01%	7.33%
89.44%				

[1]The fraction of URLs in Twiti that were detected earlier than VirusTotal.
[2]The website scan results of 2,386 URLs collected during May 1st to 14th. The clean URLs had been traced from the time of detection by Twiti to the end of May.

Volume. Twiti collected 12,765 malicious IP addresses during April. The average daily number of malicious IP addresses Twiti could collect was 426. Note that Twiti collected 16,668 IP addresses in February and 45,683 IP addresses in March. We also investigated the volume of other public IP blocklists in the same period. While public IP blocklists were mostly low in their volume, Twiti could provide a significantly large number of malicious IP addresses. Among public IP blocklists, AlienVault IP reputation showed the largest volume since it reports any malicious IPs, not limited to malware.

Exclusiveness. We judged an IP address detected by Twiti to be in VirusTotal when the IP address was flagged as malicious in VirusTotal within 30 days from its first detection date in Twiti and the IP blocklists we considered. Similarly, we checked whether an IP in Twiti was in each IP blocklist within a 30-day window. In Table 6, we provide the fraction of exclusive IP addresses to VirusTotal and each IP blocklist. Compared to VirusTotal, more than half (53.63%) of IP addresses in Twiti were exclusive. Twiti shows much higher exclusiveness (>90%) to public IP blocklists. Among public IP blocklists, Twiti shows the highest overlap (9.80%) with AlienVault IP reputation. This indicates that the contribution of each feed for IP addresses is quite unique regardless of its volume.

Latency. We defined the first detection date of a malicious IP address as the first day of its appearance in Twiti within a 30-day window. In comparison with VirusTotal, Twiti could detect 813 IP addresses 5.9 days earlier on average. Note that VirusTotal API v3.0 does not provide the detection time of a malicious IP, so we could compute the latency only for an IP that was detected first in Twiti and later in VirusTotal. We computed the difference between the first detection dates in Twiti and each blocklist within 30 days. Twiti found 274 IP 10.6 days earlier than AlienVault IP reputation, which is one of the largest public IP blocklists. Compared to other IP blocklists, Twiti could detect malicious IPs maximum 25 days earlier but their overlaps with Twiti were too small to discuss latency.

Accuracy. Unlike URLs, there is no scanning method to check whether an IP address detected by Twiti is malicious or benign. Thus, as in Li et al. [42], we only measured how many IP addresses in Twiti were in the IP allowlist constructed using top popular domains and major CDNs listed in Section 4.1. We observed that only 4 (0.03%) of IPs in Twiti are reported falsely as malicious.

4.2.4 *Domains.*

Domains were evaluated in exactly the same manner as IP addresses. The evaluation results of domains collected by Twiti during April are given in Table 7.

Volume, Exclusiveness, and Latency. Twiti collected 3,302 malicious domains during April. The average daily number of malicious domains was 110. Twiti collected 4,737 domains in February and 4,633 domains in March. Compared to VirusTotal, Twiti collected 1,888 domains (57.18%) exclusively in April. Among 1,414 domains valid for latency comparison, Twiti detected 452 domains (38.40%) 2.5 days earlier than VirusTotal and 463 domains (39.34%) on the same day.

Accuracy. Similarly to IP addresses, we only measured how many benign domains were in Twiti using Alexa, Umbrella, and Majestic

Table 6: Evaluation results of IP addresses collected by Twiti

		Volume (one month)	Exclusiveness	Overlap with Twiti	Mean Latency (days)			Accuracy			
					Twiti earlier	Same day	Twiti later	Alexa	Umbrella	Majestic	CDNs
Twiti		12,765						3	1	4	1
								4 (0.03%)			
Public Feed TI	AlienVault IP Reputation[1]	44,930	90.26%	2.78%	10.6 (44.34%)	0 (12.94%)	5.4 (42.72%)	0	0	0	2
	Bambenek_c2[1]	194	99.98%	1.03%	-	0 (100%)	-	0	0	3	2
	Feodo tracker	822	98.62%	21.41%	3.3 (16.81%)	0 (5.31%)	10.7 (77.88%)	0	0	0	0
	SSL blacklist	123	99.86%	17.07%	4.1 (60.00%)	0 (20.00%)	4.7 (20.00%)	0	0	0	1
	Mirai-related feed	6,155	99.80%	0.45%	7.2 (60.71%)	0 (10.72%)	3.8 (28.57%)	0	1	0	0
VirusTotal		-	53.63%	-	5.9 (6.37%)[2]	-[3]	-[3]				

[1]We got the written permission to use the feed for research purpose. [2]The fraction of IP addresses in Twiti that were detected earlier than VirusTotal.
[3]Not available because VirusTotal does not provide the detection time of a malicious IP.

top 25k domain lists. We observed that, in total, 2.57% of domains in Twiti were in the allowlist.

Table 7: Evaluation results of domains collected by Twiti

Volume (one month)	Exclusiveness	Mean Latency (days)		
	VirusTotal	VirusTotal		
		Twiti earlier	Same day	Twiti later
3,302	57.18%	2.5 (38.40%)[1]	0 (39.34%)	5.5 (22.26%)

Accuracy		
Alexa: 20 (0.61%)	Umbrella: 56 (1.69%)	Majestic: 29 (0.88%)
Total: 78 (2.57%)		

[1]The fraction of domains in Twiti that were detected earlier than VirusTotal.

4.3 Comparison with The Existing Systems

We compared Twiti with the existing systems that gather IOCs from Twitter: InQuest IOC DB [13] and Twitter IOC Hunter [23]. Among many other types of IOCs, we collected URLs for 2 weeks from both systems via their APIs. We checked the accuracy of the collected URLs in exactly the same manner as Twiti. Table 8 shows the evaluation results. We observed that not only can Twiti collect more URLs than both systems, but also Twiti shows a much higher accuracy than the existing systems.

Table 8: Comparison results of Twiti with the existing systems

	URL Volume	Accuracy (VirusTotal scan results)			
		Malicious	Suspicious	Clean	Not found
Twitter IOC Hunter	299	195 (62.22%)	16 (5.35%)	83 (27.76%)	5 (1.67%)
InQuest IOC DB	1,549	1,121 (72.37%)	72 (4.65%)	383 (19.56%)	53 (3.42%)
Twiti	2,386	2,134 (89.44%)	72 (3.02%)	175 (7.33%)	5 (0.21%)

5 MEASUREMENT AND ANALYSIS

5.1 The number of IOCs on Twitter

IOCs by data source. Twiti collects IOCs from tweets themselves and the links posted in tweets. Table 9 shows the data sources of Twiti and the evaluation results of IOCs in each source. Note that the exclusiveness and latency in Table 9 were computed from the comparison with VirusTotal. We observed that tweets, Pastebin.com, and AlienVault OTX Pulse were the top sources for IOC collection through Twitter – 93.26% of the collected file hashes, 94.99% of the collected URLs, 98.75% of the collected IP addresses, and 93.55% of the collected domains were obtained from those 3 data sources. Specifically, we found that **(i) Pastebin.com is the largest IOC source linked in tweets.** As observed in Table 9, 30-70% of file hashes, URLs, IP addresses, and domains in Twiti were from Pastebin.com. It also provides a large number of fresh IOCs. For example, 33.54% of the file hashes earlier than VirusTotal and 80.88% of the URLs earlier than Virustotal were shared through Pastebin.com. **(ii) Tweet texts are the largest and the most exclusive source for malicious IP collection.** We think that a short length of IPs encourages users to report IPs in tweet texts directly. Also, tweet texts are the second largest source for malicious file hashes. Except for days when a large number of file hashes were reported in tweets with Pastebin.com links, nearly 50% of file hashes were from tweet texts. Twiti could extract 60 new malicious file hashes daily from tweet texts. **(iii) AlienVault OTX Pulse is one of top IOC sources linked to tweets, but it brings a significantly large number of the delayed IOCs.** For example, 16.94% of the file hashes later than VirusTotal were from AlienVault OTX Pulse and they caused 11 days delay on average compared to VirusTotal. **(iv) URLhaus is a small source for file hashes, but it is the largest source for fresh file hashes.** 59.21% of the file hashes detected earlier than VirusTotal were reported through URLhaus links. Since URLhaus does not accepts IOCs from anonymous users, the amount is small but the quality of IOCs can be higher than other feeds that accept anonymous submissions. **(v) Security vendor blogs are the earliest source for malicious file hashes and URLs, but at the same time they are the most delayed source.** We observed that IOCs from vendors' thorough analysis reports lead to a significant delay.

IOCs by data acquisition. Twiti collects tweets by tracking both keywords and users to maximize the number of IOCs to be collected. We observed that 31.1% of the collected IOCs by Twiti were exclusively from keyword tracking, 16.3% exclusively from user tracking, and 52.6% from both methods. Interestingly, for file hashes, 95.9% of them were obtained by keyword tracking, and only 4.1% were exclusively obtained by user tracking. On the other hand, the contribution of data collection with user tracking is much larger for malicious URL, IP, and domain collection. We observed that 23.9% of the collected URLs, 38.6% of the collected IP addresses, and 31.8% of the collected domains were exclusively from user tracking.

IOCs from commercial domain. Most security vendors share a small fraction of their reports through blogs or Twitter for marketing. Security researchers often tweet or retweet such information as well. These activities make some IOC data in commercial domain move into public domain. We measured what fraction of IOCs in Twiti are from commercial domain. We considered IOCs to be from

Table 9: Data sources of Twiti and evaluation results of IOCs by data source. The exclusiveness and latency were computed from the comparison with VirusTotal. The sum of percentages in each column is slightly over 100% because of a small duplication in IOCs across data sources.

Data Source			Commercial Use	Hash				URL			
				Volume	Exclusiveness	Mean Latency (days)		Volume	Exclusiveness	Mean Latency (days)	
						Earlier	Later			Earlier	Later
Tweets (tweet texts)			Yes	7,440 (23.11%)	431 (18.58%)	1.4 (15.85%)	2.5 (17.17%)	1,245 (18.11%)	493 (20.82%)	2.1 (14.51%)	3.2 (20.20%)
External Links in Tweets	Pastebin.com		Yes	17,774 (55.20%)	1,162 (50.09%)	0.7 (33.54%)	0.9 (65.34%)	4,629 (67.35%)	1,358 (57.35%)	1.5 (80.88%)	2.7 (38.72%)
	AlienVault OTX Pulse		Yes	5,203 (16.16%)	497 (21.42%)	3.6 (3.81%)	11.0 (16.94%)	679 (9.88%)	443 (18.71%)	1.5 (1.55%)	11.0 (33.67%)
	Malware Analysis Service	URLhaus[1]	Yes	742 (2.30%)	18 (0.78%)	0.8 (59.21%)	5.2 (1.17%)	44 (0.64%)	0 (0.00%)	1.1 (0.64%)	11.5 (0.67%)
		urlscan.io[1]	with License	-	-	-	-	106 (1.54%)	58 (2.45%)	2.8 (0.87%)	5.3 (2.36%)
		VirusTotal[2]	with License	-	-	-	-	84 (1.22%)	5 (0.21%)	1.1 (0.96%)	2.3 (6.73%)
	10 major security vendor blogs		No	1,443 (4.48%)	232 (10.00%)	10.9 (0.98%)	21.6 (0.65%)	197 (2.87%)	20 (0.84%)	4.8 (1.55%)	5.7 (3.70%)

Data Source			Commercial Use	IP				Domain			
				Volume	Exclusiveness	Mean Latency (days)		Volume	Exclusiveness	Mean Latency (days)	
						Earlier	Later[3]			Earlier	Later
Tweets (tweet texts)			Yes	6,108 (47.85%)	3,368 (49.20%)	4.6 (17.71%)	-	655 (19.84%)	278 (14.72%)	3.4 (27.88%)	4.0 (31.75%)
External Links in Tweets	Pastebin.com		Yes	4,046 (31.70%)	1,829 (26.72%)	4.6 (54.12%)	-	1,245 (37.70%)	798 (42.27%)	2.1 (50.22%)	5.4 (11.07%)
	AlienVault OTX Pulse		Yes	2,459 (19.26%)	1,571 (22.95%)	9.5 (27.31%)	-	1,193 (36.13%)	791 (41.90%)	2.2 (15.71%)	5.9 (60.69%)
	Malware Analysis Service	URLhaus	Yes	-	-	-	-	-	-	-	-
		urlscan.io	with License	43 (0.34%)	37 (0.54%)	2.0 (0.12%)	-	-	-	-	-
		VirusTotal	with License	-	-	-	-	-	-	-	-
	10 major security vendor blogs		No	124 (0.97%)	54 (0.79%)	3.7 (0.74%)	-	213 (6.45%)	22 (1.17%)	2.1 (6.19%)	6.6 (7.25%)

[1]Twiti extracts URLs and their associated file hashes from URLhaus and URLs and their associated IP addresses from urlscan.io using their APIs. [2]Twiti extracts file hashes, IP addresses, and domains directly from the URL strings of virustotal.com, and URLs from VirusTotal using API. [3]Not available because VirusTotal does not provide the detection time of a malicious IP.

commercial domain if they are from accounts run by security vendors or they are from the external links corresponding to security blogs. We observed that 6% of file hashes, 5% of URLs, 1.2% of IPs, and 7.5% of domains in Twiti were from commercial domain.

IOCs by data use restriction. Twiti collects IOCs from various sources linked to tweets. Each source has different data use conditions. For example, URLhaus is licensed under CC0, which allow to use their data even commercially. By analyzing the license of each source, we found that 96% of IOCs in Twiti can be used for both non-commercial and commercial purposes, 0.4% can be used for commercial purposes with license, and 3.6% are not allowed to be used for any commercial purposes. This large portion of IOCs without data use restriction tells that Twitter is a good source for open source threat intelligence.

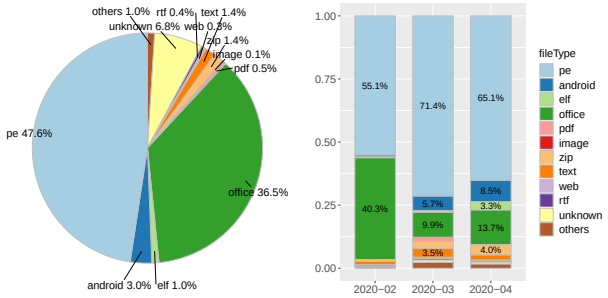

Figure 5: File type distribution of file hashes in Twiti. The "others" category includes "jar", "xml", "c++", "flash", "shell script", "wav", "avi", "iOS" files and so on.

5.2 Characteristics of IOCs on Twitter

5.2.1 File hashes.

File type. For the file hashes found in VirusTotal, we collected their file types from VirusTotal. We categorized the file types of hashes that are not found in VirusTotal as "unknown". Figure 5 shows the file type distribution of file hashes in Twiti. Although many of hashes are for PE and MS Office files, various types of malicious files are reported on Twitter, ranging from Android, Linux, iOS files to image, audio, and video files. One can acquire file hashes for a bunch of malicious Android applications as well as hashes for Linux malware from Twitter. Note that a large number of hashes for MS Office files came through Pastebin.com in early February, so Office files were dominant in that month.

Malware type. For file hashes in Twiti that were detected as malicious in VirusTotal, we analyzed their malware type using VirusTotal detection results. We picked a dominant label among different detection results of multiple antivirus engines as the malware type for a malicious file hash. Figure 6 (a) shows the malware type distribution of the file hashes detected by VirusTotal. Trojan was the most dominant threat type reported over 3 months. Except February, nearly 30% of file hashes in Twiti were ransomware. By analyzing the malware type distribution of file hashes in Twiti by file type, we observed that (i) nearly 90% of hashes for Office files were trojan-downloader, (ii) 28% of hashes for PE files were ransomware, 15% were trojan-banker, and 8% were backdoor, (iii) 30% of hashes for Android apps were trojan-banker, 17% were spyware, 12% were backdoor, and 4% were adware, and (iv) 64% of hashes for Linux malware were backdoor and 24% were trojan. For file hashes undetected by VirusTotal, we analyzed Twitter context. Figure 6 (b) shows the malware type distribution of those hashes based on Twitter context. While the majority of file hashes were shared without any malware type information, 22.6% of file hashes were mentioned with malware type. The dominant malware types of file hashes not in VirusTotal were remote access trojan (RAT) (5.5%), phishing (5.4%), and botnet (4.6%).

Malware family. We took a dominant label for malware family name after parsing antivirus detection results in VirusTotal. We

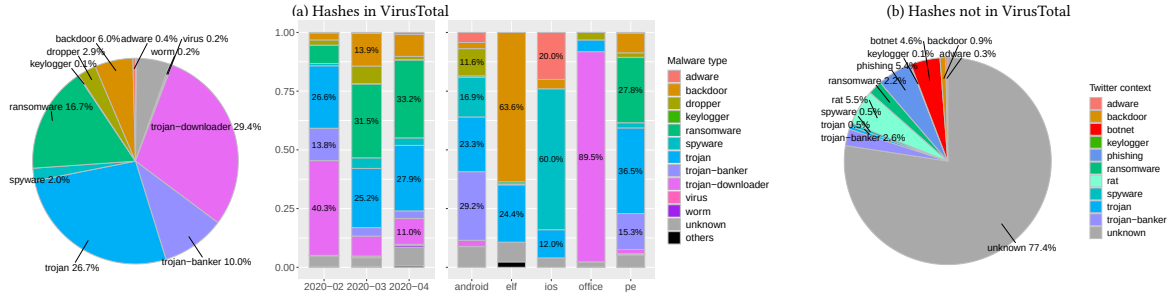

Figure 6: Malware type distribution of file hashes in Twiti for (a) hashes in VirusTotal and (b) hashes not in VirusTotal.

display top 30 malware families of file hashes in Twiti by OS in Figure 7. Emotet was the largest malware reported on Twitter, which agrees with the fact that Emotet is one of the most prevalent threats. We observed some Emotet tracking accounts on Twitter, but Emotet hashes were mostly collected by keyword tracking, which tells that various user groups report Emotet and Emotet is a serious ongoing threat. WannaCry was the second largest malware on Twitter. IoT botnets like Mirai and Gafgyt were the most dominant Linux malware on Twitter and cryptocurrency mining malware like Lady and CoinMiner was the second largest Linux malware. Banking trojans like Cerberus, Hqwar, Anubis, and Asacub were the most dominant Android malware on Twitter, and adware like HiddenAds and IconHider were the second largest Android malware. Several file hashes for Netwalker Ransomware that uses Coronavirus phishing email were reported from February 3rd to the end of April. Spyware LightRiver targeting iPhone had been mentioned from March 26th by many users for more than 2 weeks.

Early detected hashes. We analyzed file hashes detected by Twiti earlier than VirusTotal by user. There were 74 users and most of them are individual malware analysts. Top users reporting the early detected hashes are given in Table 10.

Table 10: Top users reporting file hashes earlier than VirusTotal

User	Volume	Earliness (hours)			User	Volume	Earliness (hours)		
		Mean	Min	Max			Mean	Min	Max
abuse_ch	311	17.61	0.01	74.43	pollo290987	27	51.40	7.31	640.04
malware_traffic	126	6.10	0.04	34.58	maldatabase	15	15.29	5.25	23.11
JAMESWT_MHT	78	29.36	1.20	193.55	Cryptolaemus1	14	99.53	3.50	508.47
wwp96	39	31.40	0.02	660.96	dmred1	12	153.40	0.02	447.53
lazyactivist192	30	9.62	0.01	45.29	AdamTheAnalyst	12	21.09	3.32	27.06
KanbeWorks	28	22.61	22.60	22.64	jorgemieres	8	17.06	0.46	37.49

Hashes not in VirusTotal. We analyzed who generated exclusive file hashes. There were 33 users who reported exclusive hashes more than 20 times for 3 months. 70% of them were individual malware analysts, 15% were security firms, and nearly 80% of them reported file hashes via Pastebin.com links, AlienVault OTX pulse links, malware sandbox links, or security vendor blog posts. Table 11 shows the selected top users reporting exclusive hashes.

Duration of mentions on Twitter. Figure 8 shows how many days file hashes were mentioned on Twitter. Most of file hashes had been mentioned for 1-2 days. Nearly 50% of file hashes were mentioned only a day. Meanwhile, 0.8% of file hashes had been mentioned longer than a week and, in particular, one file hash for NetWalker ransomware had been mentioned consecutively for 35 days. Malicious actors targeted the healthcare sector to take advantage of the COVID-19 pandemic, so many security experts repeatedly warned it on Twitter from early March.

Table 11: Selected top users reporting exclusive file hashes

User	Volume	Way to report
pollo290987	655	Report via Pastebin.com
Cryptolaemus1	163	Report via Pastebin.com or paste.cryptolaemus.com/
dmred1	151	Report via AlienVault OTX pulse
executemalware	139	Report via Pastebin.com
ScumBots	101	Report mostly via tweet texts
0x4A_0x4D	84	Report via AlienVault OTX pulse
IntellNetSecure	67	Report via AlienVault OTX pulse
wwp96	56	Report via tweet texts with app.any.run analysis results
virusbtn	45	Report with security vendor blog posts
HeliosCert	45	Honeypot account, Report via tweet texts with virustotal detection results

5.2.2 *URLs.*

Attack type. For the URLs in Twiti that were detected as malicious in VirusTotal, we analyzed their VirusTotal detection results [31] and observed that 75.5% of them were malware sites, 16.5% were phishing sites, and 8% were malicious sites that contain exploits or other malicious artifacts. We obtained the similar results for the URLs collected from May 1st to 15th, where 75.8% of malicious URLs were malware-related sites, 19.6% were phishing sites, and 4.6% were malicious sites. Note that 65% of phishing sites were exclusively from tweets collected by user tracking. In addition, we analyzed tweet texts since they have useful context words like "c2". We observed that 5.6% of the collected URLs co-occurred with the word "c2", which indicates that at least 5.6% of URLs in Twiti were C&C URLs. We also observed that the URLs not in VirusTotal co-occurred nearly 2 times more frequently with "c2" than those in VirusTotal, which indicates that C&C URLs usually live shortly so that VirusTotal may often fail to detect them. This tells that Twitter takes more advantage to get short-lived C&C URLs than VirusTotal.

Downloadable malware. Downloadable malware samples are particularly useful for further malware analysis. By analyzing the extensions given at the end of URLs, we observed that 32.3% of the collected URLs included downloadable file extensions such as "pdf", "zip", "exe", "apk", "sh", "jar", and "bin".

5.2.3 *Domains.*

DGA (Domain Generation Algorithm) domains. DGA domains tend to be active for short period (1-3 days). So, early detection of DGA domains is important for a blocklist to be effective. We observed that 2% of domains in Twiti were appeared with the word "dga" in tweets and they all detected a day earlier than VirusTotal. In addition, we applied a LSTM-based DGA detection algorithm [51] and observed that 5.4% of domains in Twiti were classified as DGA

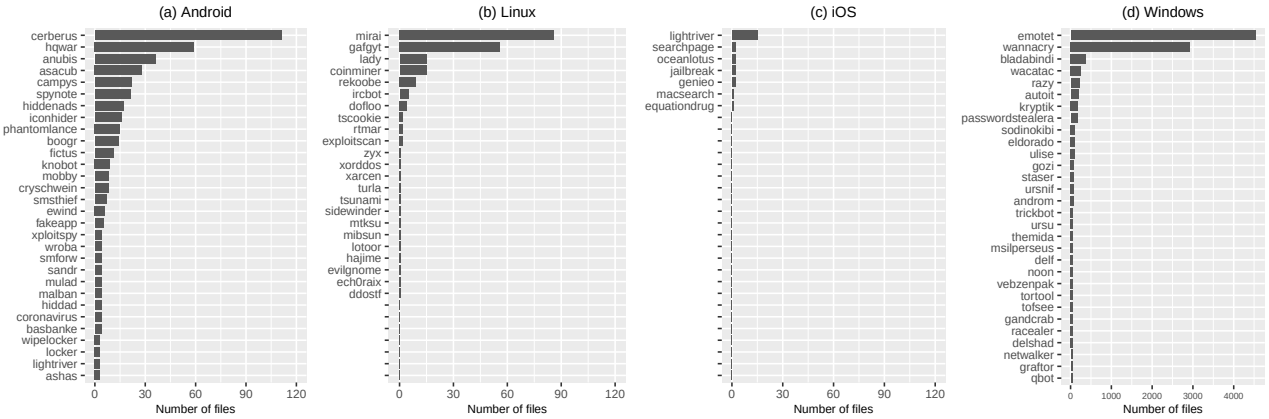

Figure 7: Top 30 malware families of file hashes in Twiti by OS platform.

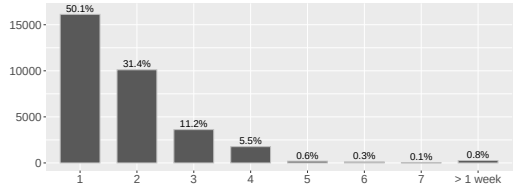

Figure 8: Distribution of the number of days file hashes have been mentioned on Twitter.

domains. Twiti detected 64% of DGA domains 1.9 days earlier than VirusTotal on average and 18% detected on the same day.

6 DISCUSSION

IOCs for other types of threats. Although we focused on malware IOCs, Twiti can be easily extended to collect IOCs for any types of attacks (e.g., phishing, spam, scanning) by adding the keywords like "phishing" and "spam" and retraining the tweet classifier.

Limitations. (1) Since Twitter is a social media platform where anyone can generate data, there are a lot of fresh threat information, but at the same time, there could be fake information. Thus, Twiti is vulnerable to data poisoning attacks, though we observe the high accuracy of Twiti in our evaluation. To overcome this weakness, one can leverage VirusTotal and IP allowlists in Section 4.2.3 to validate IOCs collected by Twiti. (2) Since Twiti collects IOCs from Pastebin.com only with a word filter, the accuracy of IOCs from it is not guaranteed when one posts some benign indicators together with malicious indicators, as we observed from the false positive analysis of URLs in Section 4.2.2. Although we observed a high accuracy (92.86% true positives and 0.03% false positives for file hashes, and 95.89% true positives and 4.1% false positives for URLs) of Twiti, its false positive rate is not low enough for use as an automatic feed. However, most public TI feeds have the limitation of a high false positive rate [50]. For this reason, public IOC feeds need a validation process before using. To reduce FPs in Twiti further, one can use Twiti as (i) an automatic feed by user selection, similar to selective pulse subscription in AlienVault OTX, and (ii) an initial source for other collaborative security systems like a multiple IP feed aggregator [47] or domain take-down system [35]. (3) Collecting IOCs from external links makes Twiti collect various types of IOCs in volume, but brings an additional dependency on external sources. Thus, for free and open source threat intelligence, Twiti cannot leverage the external sources that restrict data use.

7 RELATED WORK

IOC extraction from security blogs, forums, and mailing lists. Liao et al. [43] proposed an automatic IOC extractor from articles in technical blogs and posts in forums, known as iACE. iACE takes out sentences including IOCs from articles with a support vector machine (SVM) classifier and it then extracts IOCs from those sentences by analyzing the grammatical relations of IOCs with context terms using graph mining techniques. iACE finally converts the extracted IOCs into OpenIOC format [19]. In spite of its high accuracy, the approach does not directly apply to Twitter since grammatical rules are frequently broken on Twitter due to extensive use of non-standard language. Huang et al. [40] proposed an automatic malicious domain detector from security mailing lists. They developed a random forest classifier to identify emails discussing malicious URLs. They then derived domain names from URLs in the email text. However, treating domains of malicious URLs collectively to be malicious leads to high false positives since many recent malware attacks use legitimate services like GitHub, Google Drive, Amazon AWS, and Microsoft Azure as C&C servers [8, 10]. Gharibshah et al. [39] proposed an automatic malicious IP identifier from security forums by developing a logistic regression classifier for identifying posts reporting malicious IP addresses. Unlike iACE, Gharibshah et al. [39] and Huang et al. [40] focused on collecting lists of malicious IPs or malicious domains without any contextual information. Besides, the classifiers in both methods are designed to take the unique characteristics of data sources into account, so they are not extendable to Twitter. Recently, Zhou et al. [52] developed a deep learning-based IOC identifier from advanced persistent threat (APT) reports. They proposed a sequence labeling model based on the bidirectional long short-term memory with an attention mechanism and conditional random field layer on top to automatically locate and annotate IOC tokens such as IPs, file hashes, and URLs in sentences. However, they focused only on development of an advanced IOC extraction method and failed to provide any evaluation results on coverage and accuracy of their IOC extractor.

IOC extraction from Twitter. Dionísio et al. [38] proposed a similar deep neural network model to Zhou et al. [52], but their model

is trained with tweets to extract security-related entities from Twitter. Although their model can be extendable to extract file hashes, IPs, and URLs as in Zhou et al. [52], they focused on identification of entities for vulnerability such as company, product, product version, and Common Vulnerabilities and Exposures (CVE). Other than academic works, a number of systems have been proposed to gather IOCs from Twitter. InQuest [22] opened a tool to extract and aggregate IOCs from a plethora of open source TI feeds published through mediums such as Twitter, GitHub, and security vendor blogs. It also opened its IOC database [13] where IOCs collected since August 2019 can be searched interactively or via API. Twitter IOC Hunter [23] is a system that collects IOCs (file hashes, IP addresses, domains, URLs, emails, and CVEs) from Twitter. However, both systems simply provide lists of each IOC, but not any context information. Moreover, as shown in Section 4.3, our experiments show that the coverage and accuracy of both systems are unsatisfactory to use in practice. In particular, both systems showed considerably large false positives.

8 CONCLUSION

In this paper, we present a high-fidelity IOC extraction system for Twitter. With extensive assessment of the collected IOCs, we demonstrate that the proposed system is capable of collecting unique and accurate malware IOCs earlier than other public TI feeds. This makes Twitter a valuable open source threat intelligence feed. We also show the capability of Twitter for capturing ongoing malware attacks in volume with high accuracy and earliness. By analyzing the characteristics of IOCs in various aspects, we provide better understanding of malware IOCs on Twitter and a guide for how to leverage Twitter against malware threats.

ACKNOWLEDGMENTS

We thank VirusTotal, AlienVault, Bambenek consulting, and urlscan.io for allowing their data for this research.

REFERENCES

[1] 2019 SONICWALL CYBERTHREAT REPORT. www.sonicwall.com/lp/2019-cyber-threat-report-lp.
[2] Abuse.ch Feodo Tracker. https://feodotracker.abuse.ch/.
[3] Actionable Threat Intelligence. https://www.checkpoint.com/downloads/partners/checkpoint-intsights-solution-brief.pdf.
[4] Alexa Top 1 Million. http://s3.amazonaws.com/alexa-static/top-1m.csv.zip.
[5] AlienVault IP reputation. http://reputation.alienvault.com/reputation.data.
[6] Any.Run. https://app.any.run/.
[7] AV-TEST Security Report 2018/2019. https://www.av-test.org/fileadmin/pdf/security_report/AV-TEST_Security_Report_2018-2019.pdf.
[8] AWS, Google Cloud Popular Home for Botnet Controllers. https://www.darkreading.com/cloud/aws-google-cloud-popular-home-for-botnet-controllers/d/d-id/1330798.
[9] Cisco Umbrella 1M. http://s3-us-west-1.amazonaws.com/umbrella-static/top-1m.csv.zip.
[10] Hackers use Microsoft Azure to host malware and run C2 servers. https://www.scmagazineuk.com/hackers-use-microsoft-azure-host-malware-run-c2-servers/article/1586279.
[11] Hunting Threats on Twitter. https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/hunting-threats-on-twitter.
[12] Hybrid Analysis. https://www.hybrid-analysis.com/.
[13] InQuest Labs IOC Database. https://labs.inquest.net/iocdb.
[14] Internet Security Threat Report 2019. https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf.
[15] ioc-fanger 3.1.0. https://pypi.org/project/ioc-fanger/.
[16] iocextract 1.13.1. https://pypi.org/project/iocextract/.
[17] Majestic Million. http://downloads.majestic.com/majestic_million.csv.
[18] MalwareBazaar. https://bazaar.abuse.ch/.
[19] The OpenIOC Framework. http://www.openioc.org.
[20] OTX AlienVault. https://otx.alienvault.com/.
[21] Sources of Threat Data. https://www.recordedfuture.com/threat-data-sources/.
[22] ThreatIngestor: Extract and aggregate IOCs. https://github.com/InQuest/ThreatIngestor.
[23] Twitter IOC Hunter. http://tweettioc.com/.
[24] Twitter Search API. https://developer.twitter.com/en/docs/tweets/search/overview.
[25] Twitter Timeline API. https://developer.twitter.com/en/docs/tweets/timelines/overview.
[26] URLhaus. https://urlhaus.abuse.ch/.
[27] urlscan.io. https://urlscan.io/.
[28] Using Twitter as a source of Indicators of Compromise. https://medium.com/@cybersiftIO/using-twitter-as-a-source-of-indicators-of-compromise-bc6877fba629.
[29] The Value of Threat Intelligence: Annual Study of North American & United Kingdom Companies. https://www.anomali.com/resources/whitepapers/2019-ponemon-report-the-value-of-threat-intelligence-from-anomali.
[30] VirusTotal Contributors. https://support.virustotal.com/hc/articles/115002146809-Contributors.
[31] [n.d.]. VirusTotal Reports. https://support.virustotal.com/hc/en-us/articles/115002719069-Reports.
[32] 2019. Garmin reportedly paid multimillion-dollar ransom after suffering cyberattack. https://www.theverge.com/2020/8/4/21353842/garmin-ransomware-attack-wearables-wastedlocker-evil-corp.
[33] 2019. Security researchers take down 100,000 malware sites over the last ten months. https://www.zdnet.com/article/security-researchers-take-down-100000-malware-sites-over-the-last-ten-months/.
[34] Mitsuaki Akiyama, Takeshi Yagi, Takeshi Yada, Tatsuya Mori, and Youki Kadobayashi. 2017. Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots. *Computers & Security* 69 (2017), 155–173.
[35] Eihal Alowaisheq. 2019. Cracking wall of confinement: Understanding and analyzing malicious domain takedowns. In *The Network and Distributed System Security Symposium (NDSS)*.
[36] X. Bouwman, H. Griffioen, J. Egbers, C. Doerr, B. Klievink, and M. van Eeten. 2020. A different cup of TI? The added value of commercial threat intelligence. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association. https://www.usenix.org/conference/usenixsecurity20/presentation/bouwman
[37] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805* (2018).
[38] Nuno Dionísio, Fernando Alves, Pedro M Ferreira, and Alysson Bessani. 2019. Cyberthreat detection from twitter using deep neural networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE.
[39] Joobin Gharibshah, Tai Ching Li, Andre Castro, Konstantinos Pelechrinis, Evangelos E Papalexakis, and Michalis Faloutsos. 2017. Mining actionable information from security forums: the case of malicious IP addresses. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Springer, 193–211.
[40] Cheng Huang, Shuang Hao, Luca Invernizzi, Jiayong Liu, Yong Fang, Christopher Kruegel, and Giovanni Vigna. 2017. Gossip: Automatically identifying malicious domains from mailing list discussions. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*. ACM, 494–505.
[41] Constantinos Kolias, Georgios Kambourakis, Angelos Stavrou, and Jeffrey Voas. 2017. DDoS in the IoT: Mirai and other botnets. *Computer* 50, 7 (2017), 80–84.
[42] Vector Guo Li, Matthew Dunn, Paul Pearce, Damon McCoy, Geoffrey M Voelker, Stefan Savage, and Kirill Levchenko. 2019. Reading the Tea Leaves: A Comparative Analysis of Threat Intelligence. In *28th USENIX Security Symposium*.
[43] Xiaojing Liao, Kan Yuan, XiaoFeng Wang, Zhou Li, Luyi Xing, and Raheem Beyah. 2016. Acing the IOC game: Toward automatic discovery and analysis of open-source cyber threat intelligence. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 755–766.
[44] Edward Loper and Steven Bird. 2002. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1* (Philadelphia, Pennsylvania) *(ETMTNLP '02)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 63–70. https://doi.org/10.3115/1118108.1118117
[45] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (Baltimore, Maryland). Association for Computational Linguistics, 55–60. https://doi.org/10.3115/v1/P14-5010
[46] Niels Provos, Dean McNamee, Panayiotis Mavrommatis, Ke Wang, and Nagendra Modadugu. 2007. The Ghost In The Browser: Analysis of Web-based Malware. In *First Workshop on Hot Topics in Understanding Botnets (HotBots '07)*.
[47] Sivaramakrishnan Ramanathan, Jelena Mirkovic, and Minlan Yu. 2020. BLAG: Improving the Accuracy of Blacklists. In *Proceedings of the 27th Annual Network and Distributed Systems Security (NDSS) Symposium*.
[48] Alan Ritter, Sam Clark, Oren Etzioni, et al. 2011. Named entity recognition in tweets: an experimental study. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 1524–1534.
[49] Hyejin Shin, WooChul Shim, Jiin Moon, Jaewoo Seo, Sol Lee, and Yong H Hwang. 2020. Cybersecurity event detection with new and re-emerging words. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security (AsiaCCS)*. ACM.
[50] Sushant Sinha, Michael Bailey, and Farnam Jahanian. 2008. Shades of Grey: On the effectiveness of reputation-based "blacklists". In *2008 3rd International Conference on Malicious and Unwanted Software (MALWARE)*. IEEE, 57–64.
[51] Bin Yu, Daniel L Gray, Jie Pan, Martine De Cock, and Anderson CA Nascimento. 2017. Inline DGA detection with deep networks. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 683–692.
[52] Shengping Zhou, Zi Long, Lianzhi Tan, and Hao Guo. 2018. Automatic identification of indicators of compromise using neural-based sequence labelling. *arXiv preprint arXiv:1810.10156* (2018).