Vulnerability disclosure and cybersecurity awareness campaigns on twitter during COVID-19

Aditya Bahl  |  Aastha Sharma  |  Muhammad Rizwan Asghar

School of Computer Science, The University of Auckland, Auckland, New Zealand

Correspondence
Muhammad Rizwan Asghar, School of Computer Science, The University of Auckland, Auckland, New Zealand.
Email: r.asghar@auckland.ac.nz

Abstract
The COVID-19 pandemic introduced the new norm that changed the way we work and live. During these unprecedented times, most of the organizations expected their employees to work from home. Remote working created new opportunities for hackers since more users were making use of digital platforms for online shopping, accessing Virtual Private Network (VPN), videoconferencing platforms, and software alike. Consequently, cybercrime increased due to the increase in the attack surface, and software vulnerabilities were exploited for launching cyberattacks. There is existing research that explores vulnerability disclosure on Twitter. However, there is a lack of study on opportunistic targeted attacks where specific vulnerabilities are exploited in a way that benefit adversaries the most in times such as COVID-19. The primary aim of this work is to study the effectiveness of vulnerability disclosure pattern on Twitter in COVID-19, and discuss how Twitter can be leveraged as Open-Source Intelligence (OSINT) during a pandemic where the global users can follow a coordinated approach to share security-related information and conduct awareness campaigns. The study identifies Twitter as an apt source for conducting cybersecurity awareness campaigns as 99.83% of the security vulnerabilities are found to be accurate. The information can help global cybersecurity agencies to proactively identify vulnerabilities, coordinate activities, and plan for mitigation strategies since releasing patches from the vendor might take time.

KEYWORDS
awareness campaigns, COVID-19, cybersecurity, twitter, vulnerability disclosure

1 | INTRODUCTION

COVID-19 affected hundreds of millions of people and over 200 countries, making 2020 the most disruptive year. Due to this COVID-19 pandemic, many countries went into lockdowns that led to many employees working remotely, particularly from home.¹ The surge in online activities and software applications such as Virtual Private Network (VPN) and videoconferencing platforms led to a surge in cyberattacks due to the larger attack surface for hackers.²³ Unfortunately, adversaries exploit the flaws or weaknesses in such software to gain unauthorized access to the organizational network. These flaws or weaknesses are known as vulnerabilities, which are exploited based on their severity and have great potential for launching a successful cyberattack by hackers. Security vulnerabilities are not new, but something that existed for a long time, since 1988.⁴ Each vulnerability has a unique identifier called Common Vulnerabilities and Exposures (CVE)
ID. Such CVE identifiers help organizations/individuals to obtain details accurately and quickly. During COVID-19, working from home disclosed the extent of software vendors preparedness around the security of their products. Even after getting vulnerability fixes by the vendor, not many users run updates due to a lack of awareness.

At the beginning of 2020, 36 billion data records were exposed, which is double the number of records exposed in the whole of 2019, thus making 2020 as “worst year on record”. The Middle Eastern region also witnessed a 250% surge in cyberattacks due to the adoption of digital platforms in COVID-19. The cyberattacks ranged from data breaches to unemployment frauds and ransomware to breach of election security, impacting all in the worst possible way. Several employees working from home became vulnerable to many types of cyberattacks, one of the major being COVID-19 advisory phishing campaigns and the other being cybersecurity vulnerability exploitation. It also impacted some research communities as reported by in Attaallah et al.. In response to the cyberattacks during this period, on April 8, 2020, a joint advisory was published by the United States Department of Homeland Security (DHS), Cybersecurity & Infrastructure Security Agency (CISA), and the United Kingdom’s National Cyber Security Centre (NCSC). They discussed issues such as vulnerabilities in communications platforms, e-mail scams, malware, and included practical advice for individuals/organizations for threat vectors, Indicators of Compromise (IoC), and more. Some notable vulnerabilities exploited during December 2020 include the SolarWinds vulnerability, “CVE-2020-10148”, where SUNBURST got this name due to the trojanised component of SolarWinds Orion that affected global organizations including government agencies. The impact of such cyberattacks can be minimized by conducting better cybersecurity awareness campaigns.

Security vulnerabilities have a unique CVE-ID when they are reported and indexed in the vulnerability databases. The CVE-IDs are compiled by vulnerability databases, which provide free access to the public. National Institute of Standards and Technology (NIST) National Vulnerability Database (NVD) and MITER CVE (both sponsored by the US DHS and CISA) maintain the CVE dictionary in the public domain for free. The creation date of CVE-ID in NVD is the day when the vulnerability is indexed (it is always public); whereas, in MITER CVE, the creation date is when the CVE-ID gets reserved. Other vulnerability databases include the Packet Storm and Security Database, which covers vulnerability description, some analysis around possible fixes, mitigation, and known exploits. Since NVD and MITER CVE have strict policies before vulnerability indexing, for this study, both platforms are considered for validating the CVEs reported on Twitter to identify if any vulnerability got disclosed before its indexing. Twitter as Open-Source Intelligence (OSINT) has been an area of interest for many researchers, from tweet collection to processing, and from analyzing vulnerability lifecycle to evaluating vulnerability exploitability. Tweets are concise, more generic, and act as data storage for a lot of topics. Twitter, being a broadcast medium, is rich in cybersecurity vulnerability data in comparison to Reddit, which is a discussion forum. Reddit has vulnerability discussions initiated even before public disclosure. Twitter users also discuss vulnerabilities that are being exploited in the wild irrespective of severity. Such discussions are done in small groups of 2–13 tweets, but there was no specific account/user to follow for cybersecurity content.

The number of tweets/retweets may reflect the risk level and help find mitigation approaches. Vulnerabilities with a colloquial name attract media attention due to which the users discuss them to increase their content visibility. Twitter Dataset Collection in recent research shows different ways, but the most reliable and ethical practice is the official Twitter Application Programming Interface (API), which we choose for this study. Twitter API can be integrated with other solutions for easing accessibility to a large dataset, say fetching hourly data from Twitter and storing them in Amazon EC2, which is retrieved through Elastic Search. The other way for retrieving Twitter data is by GetOldTweets as Twitter API gives access to the past week tweets only. The main limitation of this method is that it is a way to scrape Twitter website without Twitter’s access keys. Some researchers captured the human-to-human interaction on Twitter using the existing datasets and removed Bot communications since tweets from bots are not informative. The most common way is using existing datasets, but there are no existing datasets that meet our requirements as this research requires processing the live Twitter data.

Vulnerability disclosure on Twitter is well-researched, but there is a lack of studies on opportunistic targeted attacks in which specific vulnerabilities are exploited at a time where hackers are benefitted the most. COVID-19 created such an opportunity, which led global users to work from home with software whose vendors were unprepared for such a situation.

Several security researchers, analysts, and organizations reported cybersecurity vulnerabilities, but many hackers were also exploiting such vulnerabilities. This research aims to identify who is reporting cybersecurity vulnerabilities, when, and from where via social media. Here, we leverage Twitter as OSINT platform for identifying users disclosing vulnerabilities and how cybersecurity awareness is spread on social media platforms. Twitter is chosen to carry out this study due to its unique design where the messages are limited to 280 characters, and the tweets are to the point with associating links. This study provides insights related to vulnerability disclosure and cybersecurity awareness campaigns on Twitter. Our findings will help government organizations, such as CERT, GCSB, and security agencies to conduct better cybersecurity awareness campaigns.
awareness campaigns in the future for taking proactive measures against any vulnerability exploitation. There is some
time gap between vulnerability reporting and indexing on vulnerability databases. Until then, cybersecurity campaigns
for such vulnerabilities are dispersed on social media. This time gap becomes challenging for government cybersecurity
agencies, organizations, and product vendors to develop appropriate response and protection measures. At such times,
security analysts, researchers, and stakeholders alike share their findings about such vulnerabilities. The shared solution
could be a quick fix or workaround until any official updates are available. It may also be helpful for the vendor in patching
the vulnerability; however, this is mostly seen for high and critical severity vulnerabilities. The subsequent insights on the
dataset will assist global security organizations and product vendors to better prepare and reduce exploitation impacts.

To the best of our knowledge, this study is the first of its kind that explores the effect of COVID-19 on vulnerability
disclosure and cybersecurity awareness campaigns on Twitter. Further, we identify different ways cybersecurity awareness
campaigns are conducted on Twitter, the location from which they are being conducted, and the users behind them.
Cybersecurity awareness campaigns do not have a single governing body, so a few ways are explored to identify their
legitimacy. Such work can facilitate government organizations in initiating cybersecurity awareness campaigns at a large
circle to educate the masses about persistent threats. Monitoring discussions on Twitter can also help identify zero-day
exploits and popular vulnerabilities that impact most of the users. Based on those vulnerabilities, cybersecurity agencies
can coordinate information about the latest patches, fixes, or mitigation steps.

The rest of this article is structured as follows. Section 2 reviews related work. Section 3 describes our proposed
methodology. Section 4 presents our findings and key insights. Finally, Section 5 concludes this work and provides
research directions for future work.

2 RELATED WORK

Several researchers have contributed to the application of Twitter as OSINT and vulnerability analysis. The work in this
area has evolved significantly, but the focus area has always been on the classification of security-related tweets. One
of the initial works used Twitter for developing a framework for predicting security exploits and implemented Machine
Learning (ML) techniques to provide early warnings for real-world security vulnerability exploits.\(^{16}\) The study used
Proof-of-Concept (PoC) based on publicly available exploits and reviewed several unique challenges for exploit detection.
The researchers are dependent on external or third party service providers including website blacklisting for blocking
malicious information. Subsequently, another research developed a framework called CyberTwitter based on tweets for
monitoring and exploring exploits on Twitter.\(^{20}\) Based on a keyword list, the system analyses new tweets using a Named Entity Recogniser (NER) specialized in security vulnerabilities. However, the research lacks focus on the information spread using social media intelligence for managing vulnerabilities.

Subsequent research proposes STREAMCUBE, a data structure that is based on a divide-and-conquer approach.\(^{21}\) It
aims to cluster and explore hashtags into data cubes concerning time, space, and geography. By mining geotagged tweets,
most of the reported events happening around the globe can be extracted in nearly real-time. The research also outlines
hashtags’ advantages: hashtags are less noisy and human-readable, and the semantic relation with tweets is easier to
identify. For spatial hierarchy, structures like quad-tree were used to explore data from Twitter. The authors identify that
hashtags can either include words or other hashtags that can be clustered together as per space-time hierarchy. Twitter
API was used as a data collection method, and around 9 million tweets were collected. The hashtag clusters derived from
their technique are noisy, which can affect the overall quality. There is still a need to study ways that can reduce cluster
noisiness.

Sapienza et al. developed a framework based on Twitter and the dark web forums to generate alerts that are the
early warnings of cyber threats.\(^{18}\) More specifically, they curated a framework that relies on 200 dark web forums, 69
international researchers and security analysts on Twitter. Twitter API is used to fetch hourly data from these accounts
and stored in Amazon EC2, which is retrieved through elastic search. The key phase of their framework is the warning
generation, where an alert is generated containing the threat name, frequency of words on the dark web, and Twitter.
Their results show that the method has more than 80% precision to look for early warning on Twitter and the dark web.
The main limitation of forecasting cyberattacks is that it is event-driven and dependent on unconventional signals.

Sauerwein et al. consider Twitter as a crowdsourcing platform where information about security vulnerabilities and
their patches are shared/disclosed.\(^{22}\) Using Twitter API, they collect data specifically for standard vulnerability, CVE.
They analyze the data with several phases of the vulnerability lifecycle. Also, they examine the type of CVE information
shared via tweets. For identifying bots, they used BotORNot API. They analyzed 24,000 vulnerabilities from May 2016 to
March 2018. Instead of just looking for mentions of attack, there is a need to identify all the possible indicators and events that can cause an attack. There is a need to extract indicators of compromise from Twitter to understand cyberattacks in-depth and identify the stage of attack as per Cyber Kill Chain.23

Horawalavithana et al. identify software vulnerabilities disclosure on Reddit, Twitter, and GitHub.17 They predict the software development activity on GitHub from the discussions on Twitter and Reddit. They identified publicly-known CVEs and the way discussions are happening after being disclosed publicly on platforms, such as Twitter and Reddit. Twitter is a broadcast medium rich in security vulnerability data; whereas, Reddit is a discussion forum. Vulnerability discussions on Reddit are initiated even before the public disclosure. They had a dataset of 105,596 tweets/retweets/comments having CVE-IDs discussed by 8,766 users. The dataset used for research belonged to three different datasets and had data for over a period of 1.5 years (March 2016–August 2017). For filtering out bots, a bot detector was used called Bot-hunter. The tools for identifying bots have limitations, due to which, they do not share precise results and have the known problem of false positives.24

Alves et al. explored if NIST NVD is the timeliest and richest vulnerability database or if Twitter provides timely and rich vulnerability coverage and how vulnerabilities are discussed on Twitter.13 They used a veprisk database containing information until the end of 2018, relating to many kinds of publicly available data, including all information published on NVD. They identified that NVD itself is not a complete database of vulnerabilities, and other databases (such as Packet Storm) have more entries, but all databases share information publicly after every update. Vulnerabilities mentioned on Twitter were searched on NVD and validated manually. A million tweets were manually inspected in a period of 8 months to accomplish data labelling. The final dataset had 3,461,098 tweets from early 2017 till the end of 2019. A total of 94,398 vulnerabilities were searched, and 71,850 were mentioned in tweets. Since 2010, more than 97.5% of vulnerabilities were discussed on Twitter, showing the greater coverage of CVEs in tweets. A total of 9093 vulnerabilities were checked to analyze the timeliness of discussions on Twitter. A few vulnerabilities were found on Twitter before being available on vulnerability databases. Such vulnerabilities attract attention, and the maximum of them are either high or medium severity vulnerabilities. The discussions are done in small groups of 2–13 tweets, but there was no specific account/user to follow for cybersecurity content. However, the vulnerabilities that have more than 50 replies/retweets usually have greater impacts and the discussions last 8 days.

Chandra et al. extended Endsley’s situational awareness model to build a cybersecurity awareness model.25 Their aim is to model risk-management based awareness with maturity levels to tackle cyberattacks. Risk maturity decisions for cybersecurity awareness can be taken through fuzzy Failure Mode Effect Analysis (FMEA). The decisions for maturity levels can be taken using the Capability Maturity Model (CMM) approach. The authors analyzed the data during the COVID-19 pandemic. The initial phase in the proposed model comprises the development of situational awareness, where important information assets are identified. Then, using FMEA, the risk of threats is analyzed and based on that, recommendations are predicted. The second phase includes the operations decision maturity level of cybersecurity, which determines the accuracy and priority of risk assessment. It was observed that as the maturity levels increase, there was a drop in cybersecurity incidents. A similar study on security in healthcare has been conducted using AHP-TOPSIS, which the hybrid fuzzy-based symmetrical methodology.26

Cyberattacks on technology also increased significantly, where the recent example is the brute force attack on Remote Desktop Protocol (RDP), as hackers wanted to make the most of the opportunity that they have got due to the disruption caused by COVID-19.7,27 Several mitigation strategies and guidelines have been proposed by cybersecurity agencies.28,29

### 3 PROPOSED METHODOLOGY

In this section, we present our proposed methodology used in this work, which is divided into the following stages: Understanding the State-of-the-Art, Data Collection, Data Merging, Data Cleaning, Data Validation, and Data Visualization (see Figure 1). This study is the first of its kind that explores cybersecurity awareness campaigns conducted by Twitter users around cybersecurity vulnerabilities during COVID-19. After going through the related work, we developed a thorough understanding based on which we discovered that the most suitable method for collecting data from Twitter is through the official Twitter API. The research is performed on live Twitter data collected daily between the last quarter of 2020 and the first quarter of 2021, due to which only limited data is assessed. The collected data is used for analyzing cybersecurity vulnerabilities shared by the users, including security analysts, researchers, and hackers. These users share information to increase the visibility of their content or for cybersecurity awareness. The study results have the CVE-ID of vulnerabilities, which are validated by matching it with NVD and MITER vulnerability databases. Since we report the CVE-ID of
vulnerabilities, we validate our results by checking both NVD and MITER vulnerability databases. This validation step will help us identify any fake CVE or misleading information from the dataset since Twitter is an open platform for everyone. We also explore the possibility of streamlining the validation of cybersecurity awareness campaigns on Twitter, which are generally scattered in nature.

3.1 | Data collection

We collect data using Twitter API. It provides programmatic access that allows users to create software to collect and analyze Twitter data. The free API limits access to tweets posted in the previous week. Although the free account had limitations in monthly tweet cap and data fields, we could access vulnerability relevant tweets. More specifically, the API has a limit of 30 requests per minute and 100 tweets per request; each tweet is limited to 128 characters. To find related tweets, we used “CVE-2020” because this is the standard format. For Twitter data extraction, a python library called Tweepy was used, which provides a convenient way to access the Twitter API. Tweepy has documented the API references, set of classes for all its methods representing Twitter API endpoints. Since Tweepy is a way to access Twitter API, it respects Twitter’s terms of service and is listed on the Twitter website. Tweets were collected every day from the first tweet of December 11, 2020 until the last tweet of December 31, 2020. However, the validation of CVEs was completed in the first quarter of 2021.

3.2 | Data merging

The tweets for each day were appended. All these tweets were structured into a Python Data Analysis Library (PANDAS), which provides data manipulation operations such as select, reshape, merge and various features such data wrangling and data cleaning. A data frame called Tweets_df is created in which columns show the tweet, user profile location,
tweet creation date and time, and user screen name, stored in the XLSX format. PANDAS then automatically appended all files based on similar column names, and the appended data frame is saved as an XLSX file. A total of 17,509 tweets were collected in total. The collected tweets require cleaning for further analysis due to the presence of noise or irrelevant data from the perspective of our study. Most of the data consists of missing locations or incomplete CVEs. Further, Twitter being a social media platform, some fake or false information may get included.

3.3 Data cleaning

The raw dataset consists of global users from different countries. We aim to identify who is tweeting (user), when (date and time), and from where (location) about cybersecurity vulnerabilities and awareness campaigns. An initial high-level analysis of the dataset showed that (i) the location names are not in a particular format because they include numbers (including postcodes, geo-coordinates, IP addresses, and country codes) or have alphabets in non-English languages as users can update locations as per their choice, and (ii) most locations were missing or having a NULL entry, making 60.19% of the total dataset. Based on these findings, a cleaning process was designed. Even though we chose the English language for collecting tweets, we found the location names in other languages as well. The translation of location to English is done using Googletrans, an unlimited and free python library that uses Google Translate API to make calls to methods for detection and translation of multiple entries in a single HTTP session.36

There were some locations in abbreviated forms, such as NY instead of New York. We used a python script for converting such abbreviations to full forms. The state/city names were converted to country names that made data visualization easier.

A manual check of the raw dataset shows that some locations were not picked up, such as “CALIFORNIA” and these were resolved manually. In case if users chose multiple countries as their location, the first country on the list was considered their primary one. For example, if the location is “France | USA | India”, then France is considered as the tweet country. The locations that contain numbers were resolved through online searches. For instance, both 44 145 and 91 109 are postcodes for places in Ohio and California in the USA, respectively. We also found geo-coordinates in the tweets. For example, “45.415928,−75.702755” is a location in Ottawa, Canada. Likewise, IP addresses were resolved, such as “66.66.66.66” resolves to a location in New York, USA. Further, country codes were resolved too, such as “+61” means Australia.

3.4 Data validation

NIST NVD and MITER CVE maintain the CVE dictionary in the public domain for free use37,38. Both platforms offer vulnerability feed downloads that include a short description and associated reference links from the CVE dictionary feed, as well as severity, weakness categorization, Common Vulnerability Scoring System (CVSS) scores, and vulnerable product configuration. NVD provides vulnerability feed in JSON format.37 Whereas, MITER supports XML format for each year individual dataset and cumulative dataset includes data since 1999, which is available in CSV, HTML, text, and XML formats.38 The data feed for 2020 is extracted from both platforms since almost 98.66% of CVEs in our dataset are for 2020. In our analysis, we found that only one CVE-ID is not found in both vulnerability databases and is a fake CVE. Twenty-three CVEs are found to be reserved on MITER due to which no further related information is available. Four CVEs have a disputed status, which is due to a disagreement between multiple parties’ assertion on particular security vulnerability. Three CVEs were rejected by the vendor since they are not considered as vulnerabilities. Two vulnerabilities are not supported since they are found in end-of-life products. Excluding one fake CVE, a total of 600 unique CVEs are found out of which 592 distinct CVEs were identified in 2020 and 8 distinct CVEs were disclosed before 2020.

Since all tweets are validated, a list of cybersecurity-related terms needs to be created from retweets at the next stage of preprocessing, which is useful for identifying keywords used for conducting awareness campaigns. Note that the vulnerability databases were consulted in February 2021 based on which their status is added. The 1620 short URLs collected were resolved to the expanded URLs from which the domains/subdomains were extracted. A python script is used for expanding all URLs.39 These URLs have been checked for malicious and blacklisted domains to identify the legitimacy of cybersecurity awareness campaigns.40 All links are found clean, and through a python script, the domains are extracted from the URLs. Since grouping by root domains will turn the smaller pieces of analysis into one bigger picture, which includes 92 domains.41
In this section, we discuss the insights based on the tweets we collected and analyzed. The aim of our analysis is (a) to identify who, when, and from where users tweeted about cybersecurity vulnerabilities globally, and (b) to evaluate the effectiveness of the information shared. Further, we investigate how cybersecurity awareness campaigns are conducted on Twitter. Since the collected tweets are related to CVEs, the corresponding awareness campaigns are discussed.

Figure 2 highlights countries from which the CVE-related tweets are shared the most by Twitter users. More than half (52.30%) of the Twitter dataset is contributed by users from the USA (25.06%), India (10.3%), Australia (10.22%), and the UK (6.7%). Since the USA leads with 69.3 million active users on Twitter, our dataset justifies that; this is why, we also have the most tweets from the USA. The remaining 47.70% of the total tweets is contributed by 95 countries to name a few (in the order of tweet count): France, Germany, Turkey, Spain, Canada, Italy, and Japan.

Figure 3 illustrates day-wise CVE-related tweet distribution. Because of Christmas and New Year, the tweet count decreases gradually; however, in-between, it picks up twice: (a) just before Christmas (22nd-24th December), and (b) just before New Year (27th-30th December). The first upward trend is because of CVE-2020-0986 (Windows Kernel Elevation of Privilege Vulnerability), which had a patch available in June 2020 after its disclosure in May 2020. However, in December 2020, Google hackers (Project Zero aimed at finding vulnerabilities in internal and external products) successfully exploited it and proved the patch did not fix the issue. This led the product vendor to issue a second patch in January 2021. For CVE-2020-8554 (all versions of Kubernetes API server were affected, where an attacker can intercept traffic to a cluster IP address), mitigation steps were released by the vendor and other well-known product vendors. The second upward trend is due to CVE-2020-10148 that affects “SolarWinds Orion Platform”, where authentication can be bypassed by a remote attacker for executing API commands. Attackers were actively exploiting this vulnerability. Due to the high criticality and huge impact on several organizations, many security agencies issued warnings, recommendations, as well as immediate workarounds were suggested.

There is no enough information to identify the peak of 11 December as it is the first day of our dataset, and the day with most discussions. However, the available information suggests that Microsoft released patches for 58 vulnerabilities on 08 December, and since the vulnerabilities are public, their authors were sharing the PoCs. On 14 December, Apple
released a patch for 46 security vulnerabilities in MacOS, and the number of tweets increased on the same day.\textsuperscript{49} The results show that there are two major types of discussions taking place on Twitter. Under the first type, the discussions are related to patches by the vendors. The second type is concerning cybersecurity vulnerabilities, which are disclosed by the users. The latter type attracts a lot of attention even during the holiday period. Since many organizations were affected by cybersecurity vulnerabilities, lots of attackers were seen exploiting these them and a group of users preferred to share information about them.

In our dataset, we discovered that 64.1\% of the 560 unique vulnerabilities found on Twitter either have critical or high severity, as illustrated in Figure 4. Medium priority CVEs contribute to 34.85\%, and the low is 1.05\%. Overall, 23 vulnerabilities are still under analysis at the time of vulnerability database collection (February 28, 2021), which are categorized as reserved and do not have any severity associated with them. CVSS 3.1 severity ratings are considered to create a common assessment criterion that includes the impact/privileges required/exploitation or attack complexity. The vulnerabilities discussed by more users generally have high severity and can be used by the government cybersecurity agencies as an indicator to issue a warning ahead of the product vendor. Table 1 lists down the top 10 most tweeted CVEs in our dataset, and includes tweet count, severity level, CVE-ID, and brief description. A list of the top 25 most common and dangerous weaknesses or Common Weakness Enumeration (CWE) is prepared by MITER.\textsuperscript{50} The CWE is not dependent on any product, vendor, or system, rather it depends on the vulnerability. So, 357 unique CVEs in our dataset of 560 CVEs have a CWE-ID associated with them. These CWEs make 63.75\% of CVEs in our dataset. Some of them included in the list are Cross Site Scripting (XSS), Cross Site Request Forgery (CSRF), and improper authentication.

A total of 99.83\% CVEs matched the MITER and NVD database. We validated 601 unique CVEs, out of which 592 were disclosed in 2020, eight were disclosed before 2020, and one was fake. Further analysis was carried out only for these...
TABLE 1 Top 10 most tweeted CVEs in our dataset

| Tweet count | CVE-ID          | CVE severity | Brief description                                                                 |
|-------------|-----------------|--------------|-----------------------------------------------------------------------------------|
| 193         | CVE-2020-0986   | High         | Windows kernel elevation of privilege vulnerability                                |
| 148         | CVE-2020-35 489 | Critical     | WordPress unrestricted file upload and remote code execution                        |
| 130         | CVE-2020-8554   | High         | All versions of Kubernetes API server were affected where an attacker intercepted traffic to a cluster IP address |
| 121         | CVE-2020-10 148 | Critical     | SolarWinds Orion Platform vulnerability, a remote attacker can bypass authentication for executing API commands |
| 103         | CVE-2020-16 902 | High         | Windows installer elevation of privilege vulnerability                              |
| 86          | CVE-2020-17 049 | High         | Kerberos security feature bypass vulnerability                                     |
| 69          | CVE-2020-16 875 | High         | Microsoft Exchange Server remote code execution vulnerability                      |
| 68          | CVE-2020-35 606 | High         | Any user authorized for the package updates in Webmin can execute arbitrary commands |
| 64          | CVE-2020-17 143 | High         | Microsoft Exchange information disclosure vulnerability                            |
| 63          | CVE-2020-17 140 | Medium       | Windows SMB information disclosure vulnerability                                   |

592 CVEs, which got reduced to 560 CVEs (excluding 23 reserved, four disputed, three rejected and two not supported ones), which are relevant to our study. Clearly, this indicates that the vulnerability information shared via Twitter is mostly genuine and reliable. It can also be inferred that platforms like Twitter are more popular among cybersecurity experts rather than adversaries who intend to spread fake news. In Table 1, the most discussed vulnerability is related to privilege escalation (CVE-2020-0986), making it popular among cybersecurity experts (with 193 tweets). This implies that the adversaries were aiming to get access of high-profile accounts to carry out cyberattacks. The vendors of these vulnerabilities were mainly Microsoft, WordPress, Kubernetes, SolarWinds Orion, Kerberos, and Webmin.

We can infer that out of the top 10 CVEs, two were found to be critical, and their patches are not publicly available yet. CVE-2020-35 489 is an unrestricted file upload and remote execution vulnerability in WordPress. Whereas, CVE-2020-10 148 is an authentication bypass vulnerability in SolarWinds Orion API, which an attacker can use to take remote access and execute API commands. The cybersecurity experts and professionals normally discuss critical and high severity vulnerabilities on Twitter to collaboratively discover patches and spread awareness in public.

Our study did not reflect that CVEs are disclosed on Twitter before being published on official datasets. But, it was observed that 43 CVEs were discussed on Twitter on the same day they were published at MITER CVE. We can infer that Twitter has emerged as a platform where vulnerabilities are shared quickly by experts to spread awareness, and people work together to identify fixes. As depicted in Figures 5, 94.59% of the CVEs were published on official databases, whereas 3.89% are still reserved, and their details are not published. A few CVEs (0.68%) are disputed because of disagreement between different parties. There are 0.84% of CVEs either got rejected or are unsupported for now.

Figure 6 shows the CVE-related keywords used frequently in vulnerability disclosure and cybersecurity awareness campaigns on Twitter. The word cloud depicts a few vendors/products that are most discussed since respective vulnerabilities were identified. The word cloud also shows that the vulnerability disclosure often includes vendor-specific keywords, such as Windows, SolarWinds, and Google. Whereas, cybersecurity awareness campaigns use vendor-neutral keywords, such as unpatched, mitigate, PoC, remote code, and attack.

Individuals and organizations disclose cybersecurity vulnerabilities on Twitter. Some of these are security researchers, consulting firms, security experts, and threat intelligence teams, from different parts of the world. A total of 183 distinct users were identified who disclosed vulnerabilities. Figure 7 shows the top five active users who disclose cybersecurity
vulnerabilities on Twitter. We can infer that user accounts with larger follower count tend to tweet/retweet more about a specific vulnerability. Table 2 lists down the top five users (de-identified them for preserving their privacy) who disclosed vulnerabilities for cybersecurity awareness on Twitter.

Cybersecurity awareness campaigns aim at spreading best practices that should be followed. Individuals as well as vendors make use of social media platforms to educate the masses about security flaws. Retweets are reposts of the original tweet that can help in spreading awareness. Around 63.36% of the total dataset comprised of retweets. So, we can say that most of the discussions were for spreading awareness rather than disclosing any new vulnerabilities. Security experts normally retweet to spread awareness about severe vulnerabilities or share their insight on possible fixes.
Table 2 shows that the awareness campaigns are neither conducted by an individual user nor by a single organization, but are a joint effort. Unlike CVE disclosure for which multiple databases are available, identifying the legitimacy of such campaigns is not straightforward since there is no governing body. The dispersed nature of awareness campaigns is concerning as misleading or fake campaigns can be initiated. Therefore, in this research, a manual check was performed on the user’s Twitter page to identify links to any other website such as GitHub, LinkedIn, Reddit or to check if the user/organization is verified or not. In Twitter, the verified account has a blue tick mark badge next to the name and is limited to high-profile users or organizations. We discovered only one such account in our dataset. The comments received on users post on Twitter and other platforms helped identify whether the user is genuine or not (say when the post has some negative feedback). Genuineness of the user has also been linked with their follower count and the number of tweets they post. Also, the follower count of users disclosing the vulnerabilities is very high compared to the users conducting awareness campaigns shown in Table 2, suggesting that the users who disclose vulnerabilities are more influential. To further validate the users’ influence and analyze their legitimacy, we analyzed the followers of the user accounts to check their fake followers.52 Fake followers can either be bots or inactive users; therefore, it is essential to measure user influence by using only genuine followers. We observed that the users who tweet more and are influential have fewer fake followers; whereas, less influential followers had more fake followers. And, on average, all the users had 86.52% genuine followers.

One thousand six hundred twenty short URLs included in tweets were expanded and grouped based on their root domains. The root domains can provide a high-level understanding and turn smaller pieces of information into a big picture. The cybersecurity awareness campaigns run by users who contributed 63.2% included sharing links of other posts on Twitter, linking Github code repositories, and sharing NVD links to new vulnerabilities. This suggests that tweets with these links are shared more as they are considered valid and legitimate sources. The remaining 36.8% is mostly contributed by product vendors, organizational websites/blogs, video sharing platforms, vulnerability reporting platforms, and news websites, which are difficult to validate. Although the government cybersecurity agencies conduct awareness campaigns, but it is done for a small number of vulnerabilities; whereas, private organizations were conducting the awareness regarding most vulnerabilities in their products. Mostly, individuals are involved in the process of conducting cybersecurity awareness. Therefore, government agencies need to proactively conduct awareness for most of the vulnerabilities, which will not only help small/medium organizations but also help the end users. Table 3 shows the top 10 most shared domains in CVE-related tweets.

The CVEs that are disclosed before 2020 are also found in our dataset even though the collection keyword is “CVE-2020”, which is due to their association with the recent CVEs. Mainly, the security experts discussed the older CVEs as a reference for the newly discovered ones. For instance, CVE-2019-12840 has been very popular since most of the users discussed its connection with CVE-2020-35 606. The issues currently exist because the publicly available fix for CVE-2019-12840 is still vulnerable to arbitrary command execution.

5 | CONCLUSIONS AND FUTURE WORK

The COVID-19 pandemic led to an increase in cyberattacks; the disruptions caused by it have been used as an opportunity by cyberattackers to exploit the large attack surface. The study uses Twitter to identify vulnerability disclosure and cybersecurity awareness campaigns since it contains rich and timely information about vulnerabilities. Twitter API is used to collect 17 509 tweets, which are then preprocessed using python libraries to remove irrelevant data and provide structure
to the dataset by global Twitter users. The 1620 short URLs were expanded to full URLs to derive 92 root domains. Cyber-security vulnerabilities were disclosed by individuals such as security researchers, experts, specialists, ethical hackers, and organizations such as consultancy firms, threat intelligence teams, and security assessment firms worldwide. These accounts have a huge follower count, and many users reshare their tweets. CVE-IDs were validated using MITER CVE and NIST NVD to identify the legitimacy of vulnerability discussions. The cybersecurity awareness campaigns on Twitter were dispersed and difficult to validate since there is no governing body for such activities unless the source is legitimate, say a product vendor. This raises concern because misleading campaigns can also be initiated, and the cybersecurity awareness campaigns conducted by government agencies were very limited compared to the campaigns conducted by individuals.

For this study, we manually validated awareness campaigns, including checks on website links on the users’ Twitter page (such as GitHub and LinkedIn) and checking if the user is verified on Twitter. An ML framework can be developed in the future to validate the awareness campaigns based on the replies and comments of users on such platforms. And based on the kind of feedback shared by other users, the legitimacy of awareness campaigns can be identified. The location resolving process can be automated too. ML models can be built on a large Twitter dataset, and training word embeddings would help the machine define the semantic proximity of the words.

**TABLE 3**  Top 10 domains shared on Twitter

| Domain name                     | Frequency |
|--------------------------------|-----------|
| twitter.com                    | 687       |
| github.com                     | 195       |
| nvd.nist.gov                   | 142       |
| x41-dsec.de                    | 68        |
| longterm.io                    | 34        |
| unit42.paloaltonetworks.com    | 27        |
| cve.mitre.org                  | 26        |
| bugs.chromium.org              | 19        |
| kb.cert.org                    | 18        |
| cwiki.apache.org              | 12        |

**DATA AVAILABILITY STATEMENT**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**ORCID**

Muhammad Rizwan Asghar  
https://orcid.org/0000-0002-9607-376X

**REFERENCES**

1. Parker K, Horowitz JM, Minkin R. How the Coronavirus outbreak has – and hasn’t - changed the way Americans work https://www.pewresearch.org/social-trends/2020/12/09/how-the-coronavirus-outbreak-has-and-hasnt-changed-the-way-americans-work. 2020. Accessed March 30, 2021.
2. Check Point. COVID-19 impact: Cyber criminals target zoom domains. https://blog.checkpoint.com/2020/03/30/covid-19-impact-cyber-criminals-target-zoom-domains. Accessed March 30, 2021.
3. Setera K. FBI warns of teleconferencing and online classroom hijacking during COVID-19 pandemic. https://www.fbi.gov/contact-us/field-offices/boston/news/press-releases/fbi-warns-of-teleconferencing-and-online-classroom-hijacking-during-covid-19-pandemic. 2020, Accessed March 30, 2021.
4. NIST. National vulnerability database. https://nvd.nist.gov/vuln/full-listing/1988/10. Accessed March 31, 2021.
5. Risk Based Security. Q3 report: Data breach quickview. https://pages.riskbasedsecurity.com/hubfs/Reports/2020/2020%20Q3%20Data%20Breach%20QuickView%20Report.pdf. 2020. Accessed March 31, 2021.
6. Murphy D. Middle East facing ‘cyber pandemic’ as COVID exposes security vulnerabilities, cyber chief says. https://www.cnbc.com/2020/12/06/middle-east-facing-cyber-pandemic-amid-covid-19-uae-official-says.html. 2020. Accessed March 31, 2021.
7. Lallie HS, Shepherd LA, Nurse JR, et al. Cyber security in the age of COVID-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Comput Secur. 2021;102248. https://doi.org/10.1016/j.cose.2021.102248.
8. Attaallah A, Ahmad M, Seh AH, Agrawal A, Kumar R, Khan RA. Estimating the impact of COVID-19 pandemic on the research Community in the Kingdom of Saudi Arabia. *Comput Model Eng Sci.* 2021;126(1):419-436. https://doi.org/10.32604/cmes.2021.014263
9. NCSC. Advisory: COVID-19 exploited by malicious cyber actors. https://www.ncsc.gov.uk/news/covid-19-exploited-by-cyber-actors-advisory. 2020. Accessed March 31, 2021.
10. Malwarebytes Labs. Backdoor. Sunburst https://blog.malwarebytes.com/detections/backdoor-sunburst. Accessed: March 31, 2021.
11. NIST. National vulnerability database. https://nvd.nist.gov. Accessed March 31, 2021.
12. MITRE. CVE. https://cve.mitre.org. Accessed March 31, 2021.
13. Alves F, Andongabo A, Gashi I, Ferreira PM, Bessani A. Follow the blue bird: a study on threat data published on twitter. In: Liqun Chen, Ninghui Li, Kaitai Liang, Steve A Schneider, eds. *European Symposium on Research in Security (USENIX Security 15).* USENIX Association; 2015: 1041–1056.
14. PacketStorm. https://packetstormsecurity.com. Accessed April 9, 2021.
15. Security Database. https://www.security-database.com. Accessed April 9, 2021.
16. Sabottke C, Suciu O, Dumitras, T. Vulnerability disclosure in the age of social media: exploiting twitter for predicting real-world exploits. In: Jaeyeon Jung, Thorsten Holz, eds. *24th USENIX Security Symposium (USENIX Security 15).* USENIX Association; 2015: 1041–1056.
17. Horavalithana S, Bhattacharjee A, Liu R, Choudhury N, O. Hall L, Iamnitchi A. Mentions of security vulnerabilities on Reddit, Twitter and GitHub. In *IEEE/WIC/ACM International Conference on Web Intelligence,* 2019: 200–207. https://doi.org/10.1145/3350546.3352519.
18. Sapienza A, Bessi A, Damodaran S, Shakarian P, Lerman K, Ferrara E. Early warnings of cyber threats in online discussions. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW),* IEEE; 2017: 667–674. https://doi.org/10.1109/ICDMW.2017.94.
19. Shrestha P, Sathanur A, Maharjan S, Saldanha E, Arendt D, Volkova S. Multiple social platforms reveal actionable signals for software vulnerability awareness: a study of GitHub, Twitter and Reddit. *Plos One.* 2020;15(3):e0230250.
20. Mittal S, Das PK, Mulvad V, Joshi A, Finin T. CyberTwitter: using twitter to generate alerts for cybersecurity threats and vulnerabilities. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* IEEE; 2016: 860–867. https://doi.org/10.1109/ASONAM.2016.7752338.
21. Feng W, Zhang C, Zhang W, et al. STREAMCUBE: hierarchical spatio-temporal hashtag clustering for event exploration over the twitter stream. In *2015 IEEE 31st International Conference on Data Engineering,* IEEE; 2015: 1561–1572. https://doi.org/10.1109/ICDE.2015.7113425.
22. Sauerwein C, Sillaber C, Huber MM, Mussmann A, Breu R. The tweet advantage: an empirical analysis of 0-day vulnerability information shared on twitter. In *IFIP International Conference on ICT Systems Security and Privacy Protection,* Springer, Cham; 2018: 201–215. https://doi.org/10.1007/978-3-319-99828-2_15.
23. Lockheed Martin. The Cyber Kill Chain. https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html. Accessed March 30, 2021.
24. Rauchfleisch A, Kaiser J. The false positive problem of automatic bot detection in social science research. *PloS One.* 2020;15(10):e0241045.
25. Chandra NA, Ratna AAP, Ramli K. Development of a cyber-situational awareness model of risk maturity using fuzzy FMEA. In *2020 International Workshop on Big Data and Information Security (IWBUS),* IEEE; 2020: 127–136. https://doi.org/10.1109/IWBIS50925.2020.9255543.
26. Kumar R, Pandey AK, Baz A, et al. Fuzzy-based symmetrical multi-criteria decision-making procedure for evaluating the impact of harmful factors of healthcare information security. *Symmetry.* 2020;12(4):664.
27. Galov D. Remote spring: the rise of RDP bruteforce attacks. https://securelist.com/remote-spring-the-rise-of-rdp-bruteforce-attacks/96820. 2020. Accessed April 1, 2021.
28. NCSC. Working remotely: advice for organisations and staff. https://www.ncsc.govt.nz/newsroom/working-remotely-advice-for-organisatios-and-staff. 2020. Accessed April 1, 2021.
29. CERT NZ. Enabling staff to work remotely https://www.cert.govt.nz/business/guides/working-remotely-securely. Accessed April 1, 2021.
30. Twitter. Pricing – Twitter developers. https://developer.twitter.com/en/pricing/search-fullarchive. Accessed April 1, 2021.
31. What is the CVE ID syntax and when did it change? https://cve.mitre.org/about/faqs.html#cve_id_syntax_change. Accessed April 16, 2021.
44. CVE-2020-8554. Man in the middle vulnerability in kubernetes – top recommendations. https://securityboulevard.com/2020/12/cve-2020-8554-man-in-the-middle-vulnerability-in-kubernetes-top-recommendations. Accessed April 7, 2021.
45. Multiple vulnerabilities in solarwinds orion could allow for arbitrary code execution. https://www.cisecurity.org/advisory/multiple-vulnerabilities-in-solarwinds-orion-could-allow-for-arbitrary-code-execution_2020-166. Accessed April 7, 2021.
46. The solarwinds cyber-attack: What you need to know. https://www.cisecurity.org/solarwinds. Accessed April 7, 2021.
47. Microsoft. December 2020 security updates. https://msrc.microsoft.com/update-guide/releaseNote/2020-Dec. Accessed April 7, 2021.
48. Microsoft December 2020 patch tuesday fixes 58 vulnerabilities. https://www.zdnet.com/article/microsoft-december-2020-patch-tuesday-fixes-58-vulnerabilities. Accessed April 7, 2021.
49. Apple releases MacOS big sur 11.1 with AirPods max support and Mac App Store privacy labels. https://www.macrumors.com/2020/12/14/apple-releases-macos-big-sour-11-1. Accessed April 7, 2021.
50. 2020 CWE Top 25 most dangerous software weaknesses. https://cwe.mitre.org/top25/archive/2020/2020_cwe_top25.html. Accessed April 7, 2021.
51. CVE details. https://www.cvedetails.com/top-50-products.php?year=2019. Accessed April 7, 2021.
52. Fake followers audit. https://sparktoro.com/fake-followers. Accessed April 7, 2021.

How to cite this article: Bahl A, Sharma A, Asghar MR. Vulnerability disclosure and cybersecurity awareness campaigns on twitter during COVID-19. Security and Privacy. 2021;e180. https://doi.org/10.1002/spy2.180