The Tweet Advantage: An Empirical Analysis of 0-Day Vulnerability Information Shared on Twitter

Clemens Sauerwein, Christian Sillaber, Michael M. Huber, Andrea Mussmann, Ruth Breu

The Tweet Advantage: An Empirical Analysis of 0-Day Vulnerability Information Shared on Twitter

Clemens Sauerwein ✉, Christian Sillaber, Michael M. Huber, Andrea Mussmann, and Ruth Breu

University of Innsbruck, Department of Computer Science, Technikerstraße 21a, A-6020 Innsbruck, Austria
Clemens.Sauerwein@uibk.ac.at
https://www.uibk.ac.at/informatik/

Abstract. In the last couple of years, the number of software vulnerabilities and corresponding incidents increased significantly. In order to stay up-to-date about these new emerging threats, organizations have demonstrated an increased willingness to exchange information and knowledge about vulnerabilities, threats, incidents and countermeasures. Apart from dedicated sharing platforms or databases, information on vulnerabilities is frequently shared on Twitter and other social media platforms. So far, little is known about the obtainable time advantage of vulnerability information shared on social media platforms. To close this gap, we identified 709,880 relevant Tweets and subsequently analyzed them. We found that information with high relevance for affected organizations is shared on Twitter often long before any official announcement or patch has been made available by vendors. Twitter is used as a crowdsourcing platform by security experts aggregating vulnerability information and referencing a multitude of public available webpages in their Tweets. Vulnerability information shared on Twitter can improve organizations reaction to newly discovered vulnerabilities and therefore help mitigating threats.

Keywords: Information Security · Shared Cyber Security Information · Social Networks · Data Mining · Twitter · Security Incidents

1 Introduction

In recent years, cyber attacks have increased significantly in number and have become more sophisticated while the time frame for organizations to react shrinks constantly [16]. To counteract new threats, organizations are implementing vulnerability management processes, increase the internal dissemination of security information and conduct awareness trainings as an integral part of their security management process [30].

The time it takes to develop, distribute and implement a patch that adequately fixes a vulnerability creates a window of exposure that organizations seek

to minimize [2, 13]. Organizations need to be able to quickly react on newly occurring threats. In doing so, they need up-to-date information about vulnerabilities, exploits, incidents, and available countermeasures [21]. Threat Intelligence Sharing Platforms or vulnerability databases which keep track of security related issues for different software applications are potential information sources supporting these activities [27]. For example, the National Vulnerability Database[1] holds more than 104,140 Common Vulnerability Exposures (CVEs), which are standardized descriptions for publicly known information security vulnerabilities and exposures [20].

Vulnerability information is also discussed and shared on social media platforms and related informal channels [26]. Moreover, there are grounds for the assumption that vulnerabilities may be discussed on Twitter before public disclosure [34]. However, research and practice lacks an empirical analysis of this assumption. The research at hand addresses this gap by extracting Tweets containing vulnerability information and analyzes if there is a time advantage of obtaining vulnerability information from Twitter compared to conventional sources. Moreover, this contribution extends previous research (cf. [34]) by analyzing the contents of referenced webpages in Tweets regarding vulnerabilities.

Our data collection is based on CVEs which are standardized security vulnerability identifiers [20]. In total, we collected a set of 709,880 Tweets between May 23, 2016 and March 27, 2018. Based on this dataset, the paper provides a comprehensive analysis of the collected information. We analyze how the collected vulnerability information maps to the different phases of the vulnerability lifecycle [14]. We also assess whether vulnerabilities are discussed on Twitter before the vulnerabilities official public disclosure. Moreover, we briefly examine what types of vulnerability information (e.g. descriptions of vulnerabilities, demonstrations of exploits,...) are referenced on Twitter.

The remainder of this work is structured as follows. Section 2 discusses related work regarding information security research based on Twitter data, and background information regarding the vulnerability lifecycle and assignment of CVE identifiers. Section 3 describes our research methodology, including data collection, processing and analysis. Section 4 provides an analysis of the collected Tweets. Section 5 discusses the results and limitations of the research at hand. Finally, Section 6 concludes the paper and provides outlook on future research.

2 Related Work and Background Information

In this section we discuss related work, and provide background information regarding the vulnerability lifecycle model and assignment of CVE identifiers.

2.1 Related Work

Analysis of data obtained from Twitter has been used successfully in a wide variety of different research applications, ranging from earthquake detection [9]

[1] https://nvd.nist.gov/ (Accessed: May 30th, 2018)

and epidemiology [31], to stock market analysis [5] or identification of cyber-bullying [1].

Since 8% of all hyperlinks shared on Twitter are phishing or spam [15], one field of research in the context of Twitter focuses on spam abuse detection and prevention: Spam detection research on Twitter ranges from general [3, 36] to political spam abuse detection [7, 23]. Moreover, the detection of accounts abused for spam [19, 29], and spamming strategies [8] have been intensively analyzed.

Other works in the field focus on the detection of security information shared on Twitter. For example, Erkal et al. applied machine learning to Twitter data to distinguish between cyber security and non-cyber security related Tweets [12]. Several authors introduced approaches that can automatically detect cyber security events on Twitter [10, 18, 24, 28, 35]. Sabottke et al. [25] introduced an approach to predict exploits based on Twitter discussions. Syed [33] analyzes the impact of social media posts on the patching behavior of vendors.

Apart from spam abuse and security information detection, Twitter was used as an information source for empirical investigations in the field of information security. For example, Jeske et al. [17] examined the extent to which specific communities of Twitter users were engaged in the debate about the Heartbleed security bug. Moreover, Syed et al. [34] conducted an empirical study to identify the major content categories contained in vulnerability discussions on Twitter and what factors impact the Re-tweeting of these contents. Moreover, Bilge and Dumitras [4] empirically analyze data from real hosts to identify attacks before and after public disclosure.

Our contribution extends the discussed related work, especially [34], through an empirical analysis of the obtainable time advantage from vulnerability information shared on Twitter. In doing so, we analyze the available information with respect to the different phases of the vulnerability lifecycle [2, 14] and the types of referenced information. To the best of our knowledge, no prior empirical research has been conducted to analyze the time behavior of vulnerability related Tweets on Twitter.

2.2 Vulnerability Lifecycle

As depicted in Figure 1, the lifecycle of a vulnerability [2] can be divided into three phases: The (1) *black risk phase*, (2) *gray risk phase*, and (3) *white risk phase*.

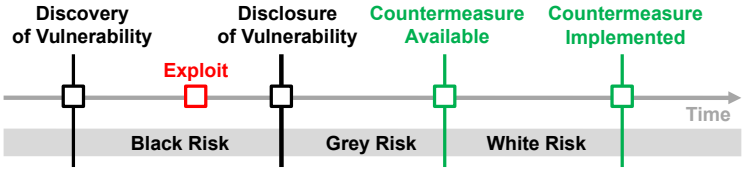

Fig. 1. Vulnerability lifecycle (based on [2, 14])

A vulnerability is in the (1) *black risk phase* from the time of its discovery to the time of its validated public disclosure to a wider audience. In this context, it is worth mentioning that the discovery of a new vulnerability is not publicly known until public disclosure. Since it is not feasible to read all security related information sources (e.g., mailing lists) or underground information sources (e.g., Darknet) to identify new vulnerabilities we follow Frei et al. [14] who define the time of public disclosure "*[...] as the first date a vulnerability is described on an information channel where the disclosed information on the vulnerability is (a) freely available to the public, (b) published by a trusted and independent channel and (c) has undergone analysis by experts such that risk rating information is included*" [14]. Accordingly we assume that the public disclosure of a vulnerability is the publication of a vulnerability on the National Vulnerability Database[2].

The *black risk phase* is followed by the (2) *gray risk phase*. The *gray risk phase* begins with the public disclosure or 0-day of the vulnerability and ends with the availability of a vendor approved countermeasure (e.g., patch). Frei et al. describe this timespan as the window of exposure. Risk exposure in this phase is especially high as the public is aware of a vulnerability without official countermeasures being available [14].

The (3) *white risk phase* completes the lifecycle. It starts with the release of a vendor approved countermeasure (e.g., patch etc.) and concludes with its roll out on the vulnerable system. Unfortunately, in most of the cases, the availability of official countermeasures lags behind the public disclosure of a vulnerability [14].

As depicted in Figure 1, an exploit might become available in any phase. According to [14], the point in time during the vulnerability life cycle when an exploit appears has immense impact on the risk for the affected system. For example, as depicted in Figure 1, if an exploit appears during the *black risk phase*, nobody is yet aware of the vulnerability, though it can be exploited.

2.3 Assignment of CVE Identifiers

The assignment of CVE identifiers takes place before official public disclosure (during the *black risk phase*) of a vulnerability on the National Vulnerability Database [20]. CVE identifiers enable researchers and vendors to assign IDs to new vulnerabilities and facilitate the tracking of vulnerabilities over time on different information channels [20]. The general workflow of assigning CVE identifiers is the following: First, a researcher or vendor detects a new vulnerability. Second, she requests a CVE identifier from a CVE Numbering Authority which assigns it to the detected vulnerability. From this moment on, the vulnerability appears as a candidate on the MITRE webpage[3] and all discussion regarding this vulnerability should contain the assigned CVE identifier. In the meantime security experts from the MITRE or National Vulnerability Database analyze and validate the vulnerability to become an official approved CVE entry. Thereby, the security experts decide if it is a vulnerability or not. For example, they eliminate

[2] https://nvd.nist.gov/ (Accessed: May 30th, 2018)
[3] https://cve.mitre.org/cve/ (Accessed: May 30th, 2018)

false positives or duplicates. If it is a validated CVE entry, it will be published on the National Vulnerability Database and MITRE marks it as validated. This point in time can be described as the public disclosure of a vulnerability on a trusted, independent and publicly available channel (cf. [14]).

3 Research Methodology

In order to facilitate the efficient collection of Tweets distributed over different accounts, we collected data using a keyword search through the Twitter Streaming Application Programming Interface (API). We searched for Tweets matching the aforementioned CVE identifiers. CVE identifiers are unique and correspond to the following pattern: `CVE-\d{4}-\d{4}\d*` [20] (e.g., a valid CVE identifier might be CVE-2016-5696). In doing so, we ensured that we only obtain Tweets containing vulnerability related information. To eliminate Tweets containing wrong or untrustworthy information, we cross-validated CVE identifiers included in the obtained Tweets against CVE identifiers listed on the MITREs webpage (assigned CVEs) and National Vulnerability Database (expert validated CVEs). Tweets without any matches were excluded from further processing.

We enriched the obtained Tweets with additional information from the National Vulnerability Database containing more than 104,140 (Accessed: May 30th, 2018) validated CVE entries with corresponding descriptions. The resulting dataset for statistical analysis contained the following information about every collected Tweet: Release date, content of the Tweet, user name, referenced websites, retweet status, and further information about the considered vulnerability including the date of public disclosure, description of vulnerability, and associated CVE identifiers. We converted all timestamps (e.g., release dates) to Central European Time Zone (CET). Moreover, if available, we enriched the collected Tweets with information about known exploits obtained from the Exploit Database[4], including descriptions and release dates of exploits for certain vulnerabilities. In order to identify and label Tweets created by bots, we used the BotOrNot API [11] which implements a supervised learning method for identifying social bots. We selected the BotOrNot API due to the high accuracy it has shown in previous research [32].

We used R-project[5] to statistically analyze the collected and processed data. In doing so, we mapped the collected Tweets to the vulnerability lifecycle model of each identified CVE and calculated the timespan between the occurrence of the first Tweet regarding a CVE and its public disclosure on the National Vulnerability Database. Secondly, we analyzed the referenced websites in the Tweets and manually classified them according to the four types defined in Section 4.3. In doing so, two of the authors of this publication independently classified the 500 most frequently referenced websites. Finally, the two classification results were compared. If discrepancies were identified they were resolved through discussion and reclassification by the two authors.

[4] https://www.exploit-db.com/ (Accessed: May 30th, 2018)
[5] https://www.r-project.org/ (Accessed: May 30th, 2018)

4 Results

The following section outlines the results of our data analysis and discusses them. At first, we describe some general observations made about the collected Tweets. Secondly, we map the collected Tweets to the vulnerability lifecycle model [2]. Finally, we briefly analyze and discuss the contents of the collected Tweets.

4.1 General Observations

The described data collection method delivered a set of 709,880 Tweets, containing 24,267 distinct CVE identifiers over a period of one year, ten months and five days, starting on May 23, 2016 and ending on March 27, 2018.

The raw dataset contains 205,255 Re-tweets[6], which accounts for 28.9% of the collected Tweets. On average, 1,077 Tweets appeared per day, where the minimum was 24 and the maximum was 5,531 Tweets per day. Moreover, per CVE on average 29 Tweets appeared with a minimum of one tweet per CVE and a maximum of 10,700 Tweets per CVE. The obtained Tweets were generated by 58,644 different user accounts.

While the average number of Tweets per user is 12.1 Tweets, the maximum amount of Tweets for a user in our data set is 36,890. Moreover, 70% of the Tweets in our dataset were generated by 100 accounts, 92% of which are bots. In total, 82% of the Tweets were generated by bots and 8% by human users. The remaining 10% could not be classified by the BotOrNot API. An analysis of the top 100 user time zones showed that users from the United States of America and Europe dominated our dataset, accounting for roughly 90% of all Tweets.

4.2 Mapping the Collected Tweets to the Vulnerability Lifecycle

In order to analyze if there is a time advantage of obtaining vulnerability information from Twitter compared to conventional sources (i.e. official Vulnerability databases), we mapped the collected Tweets to the vulnerability lifecycle model [2] (as described in Section 2). We created a timeline for every CVE identifier containing all Tweets referencing it, its public disclosure date, and (if available) the release date of the first exploit. As discussed in Section 3.1, we assume the publication of a vulnerability on the National Vulnerability Database as the point of public disclosure. As mentioned in Section 3, it is worth mentioning that a differentiation between *gray* and *white risk phase* was out of scope for this study as we did not include information on countermeasures.

In order to better understand how information is shared on Twitter during *black risk* and *gray* or *white risk phases*, we define and distinguish between the following three patterns: (1) *Tweet-Disclosure-Pattern*: The first tweet referencing a vulnerability appears before its public disclosure during the black risk phase. (2) *Disclosure-Tweet-Pattern*: The first tweet referencing a vulnerability

[6] A retweet is a repost of a message posted by an user.

appears after its public disclosure during the *gray/white risk phase*. (3) *Tweet-Pattern*: A tweet referencing a vulnerability appears but public disclosure did not take place during our observation.

Our investigations covered 24,267 different vulnerabilities in total. Our analysis identified the *Disclosure-Tweet-Pattern* for 69.6% (16,879), the *Tweet-Disclosure-Pattern* for 25.7% (6,232) and the *Tweet-Pattern* for 4.7% (1,156) of the vulnerabilities. This result shows that the majority of vulnerabilities receive increased attention during the *gray/white risk phase* of the vulnerability lifecycle, while approximately one quarter of the vulnerabilities is discussed on Twitter during the *black risk phase*.

The *Tweet-Disclosure-Pattern* is interesting for organizations and security experts as information about vulnerabilities is shared before the public disclosure of the vulnerability to a wider audience. Consequently, there might be a time advantage of obtaining vulnerability information from Twitter compared to conventional sources, such as the National Vulnerability Database.

In order to get a better understanding we want to take a closer look on CVE-2016-5696, where the *Tweet-Disclosure-Pattern* can be observed. CVE-2016-5696 describes a vulnerability in the Linux kernel that allows the hijacking of TCP connections in a fast and reliable way [6]. The vulnerability was officially published on the National Vulnerability Database on August 6, 2016, which marks the day of public disclosure. A corresponding patch for the Red Hat enterprise Linux Kernel was released[7] on August 18, 2016 which marks the beginning of the *white risk phase*. Figure 2 shows the number of Tweets collected per day over time referencing the CVE-2016-5696 vulnerability. The timeline depicts a 45 day period beginning with the first occurrence of a relevant tweet on the July 12, 2016 and ending on August 25, 2016 [8]. The public disclosure and patch release dates are marked in Figure 2. Accordingly, the timeline can be divided into the three phases of the vulnerability lifecycle, which are shown on the X-axis. During the black risk phase, 14 Tweets were collected, the first of which was tweeted on July 12, 2016. These Tweets contained information regarding the vulnerability by referencing blogs of security experts discussing them. As mentioned above and depicted in Figure 2, the *gray risk phase* was initialized by the public disclosure of the vulnerability on August 6 ,2016 to a wider audience by the National Vulnerability Database. It is significant that the amount of daily Tweets increased immediately after public disclosure. In total, 741 Tweets were collected during the *gray risk phase*. These Tweets contained demonstration videos on how to exploit the vulnerability, descriptions of countermeasures, and discussions regarding the vulnerability. On August 18, 2016 the patch for CVE-2016-5696 was released, which concluded the *gray risk phase* and started the white risk phase. During this phase, we observed many Tweets about the patch and its availability.

The most interesting observation about the aforementioned *Tweet-Disclosure-Pattern* is that there are Tweets available discussing the vulnerability during the

[7] https://rhn.redhat.com/errata/RHSA-2016-1633.html (Accessed: May 30th, 2018)
[8] Note: Due to space limitations, we did not show the full white risk phase.

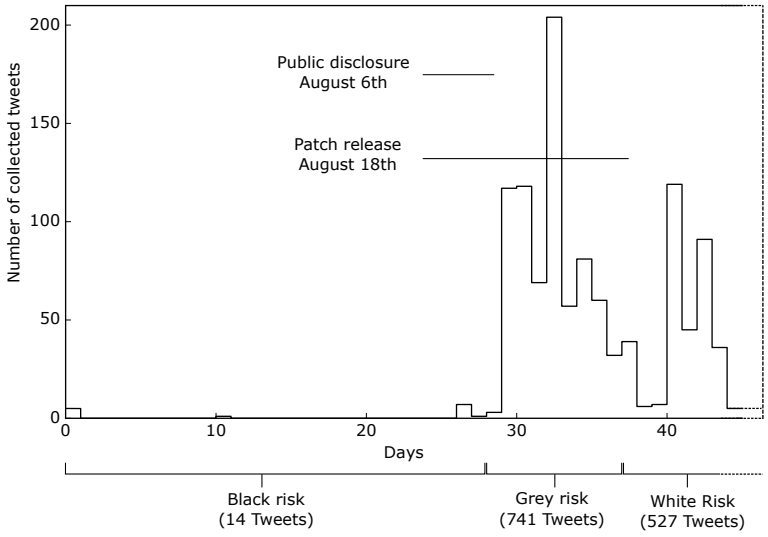

Fig. 2. The vulnerability lifecycle of CVE-2016-5696

black risk phase – i.e. before public disclosure. Our investigations showed that minimal one, maximal 972 and on average 9.66 Tweets per CVE appeared on Twitter before the public disclosure of a vulnerability. Moreover, we observed that the first Tweet of a vulnerability covering the *Tweet-Disclosure-Pattern* appears on average 33.10 days before public disclosure.

Figure 3 shows the frequency distribution of Tweets regarding vulnerabilities following the *Tweet-Disclosure-Pattern*. The Y-axis describes the number of Tweets and the X-axis the timeline, where 0 marks the time of public disclosure. For example, a tweet appearing at a point marked with -10 means that it was tweeted ten days before public disclosure. Moreover, the red bars show the total number of Tweets excluding Re-tweets and the blue bars show the number of Re-tweets. As depicted in Figure 3, an information peak can be identified in proximity to the public disclosure (point in time = 0) where in total 5,182 Tweets appeared. Moreover, a second peak can be observed six days before public disclosure. A closer look on the Tweets appearing on this day showed that high profile vulnerabilities gained increased attention. For example, CVE-2017-0144[9] which is used by the Wannacry Ransomware [22] was discussed intensively. Figure 3 shows that there is an increased attention during the black risk phase, starting six to seven days before public disclosure, increasing until public disclosure and rapidly decreasing after it. Moreover, the retweet rate is higher before public disclosure compared to the time afterwards.

A comprehensive analysis of the whole dataset, including all three (*Tweet*, *Tweet-Disclosure*, *Disclosure-Tweet*) patterns showed that the first tweet of a vulnerability appears on average 4.2 hours after public disclosure. Moreover, on

[9] https://nvd.nist.gov/vuln/detail/CVE-2017-0144 (Accessed: May 30th, 2018)

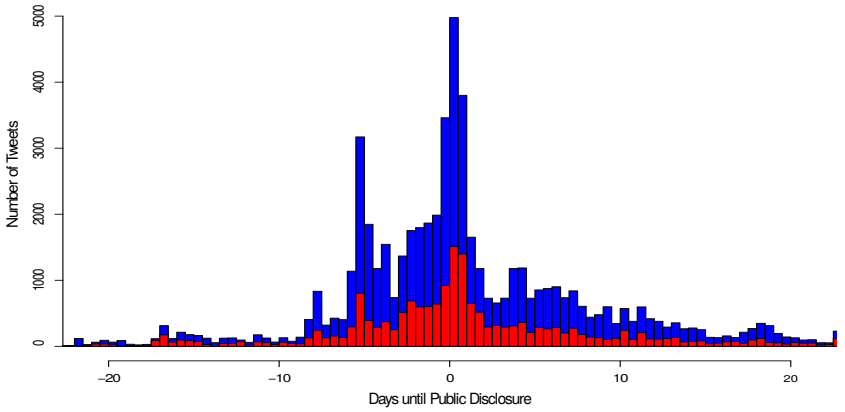

Fig. 3. Frequency distribution of 100 most mentioned vulnerabilities following the *Tweet-Disclosure-Pattern*.

average per vulnerability, 2.7 Tweets appear before and 18.68 Tweets after its public disclosure.

4.3 Contents of the Collected Tweets

The analysis of the contents of the collected Tweets showed that in 80% (567,904 Tweets) of the cases, security experts and bots tend to reference external websites in their posts. This might be traced back to a lack of space in Twitter posts. Previous research [34] has analyzed the contents of Tweets discussing vulnerability information on Twitter without considering the referenced websites. We extend it through analyzing the contents of the referenced websites with respect to the vulnerability lifecycle model. Our analysis of the 500 most referenced websites showed that they primarily contain the following four types of information:

1. *Descriptions of vulnerabilities*: Researchers, vendors, security experts as well as bots often reference websites describing emerging vulnerabilities in more detail. For example, we identified Tweets referencing security mailing lists, expert blogs or vulnerability databases. Moreover, we found that this type of information is quite frequent, accounting for 90% of all Tweets during black risk phase.
2. *Demonstrations of exploits*: Twitter users inform the Twitter community of the existence of an exploit for a certain vulnerability by referencing videos or blog entries demonstrating the exploit. For example, our data dump contains 1,603 Youtube videos describing how to exploit various vulnerabilities. We observed that this type of information can be identified on Twitter throughout all phases of the vulnerability lifecycle.
3. *Unofficial proposals of countermeasures*: Normal Twitter users frequently propose links to unofficial work arounds and countermeasures in order to

mitigate the severity of a vulnerability prior to official patch release. For example, we identified Tweets discussing and referencing security blogs describing unapproved countermeasures. This type of information primarily can be found during gray risk phase and is of potential value since an official countermeasure is not available.

4. *Announcements of patch releases*: Vendors inform the community about official available countermeasures (e.g., patches) by referencing their webpages for further information regarding it. It is not surprising that this type of information appears on Twitter in close proximity to the time an official patch is released.

The identified four types of information are distributed as follows: 59.3% *Descriptions of vulnerabilities*, 12.2% *Announcements of patch release*, 13.5% *Unofficial proposals of countermeasures*, 6.7% *Demonstrations of exploits*. The remaining 8.3% were not classifiable.

5 Discussion and Limitations

In the following Section, we discuss the results and their implications for research and conclude this Section with a discussion of the limitations of the research at hand.

5.1 Discussion of Results

The main goal of our investigations was to analyze the nature, timeliness and types of vulnerability information shared on Twitter in order to show how organizations can benefit. Vulnerability information shared on Twitter should be treated with care as it originates from not validated sources. In order to counteract this limitation, we based our data collection on CVE identifiers (see Section 3.2) and cross-validated the collected Tweets with information obtained from the National Vulnerability Database. In doing so, we ensured that we collected information covering information security topics, like discussions on vulnerabilities, threats or countermeasures.

The subsequent mapping of the collected Tweets to the vulnerability lifecycle model [14] showed that the majority of vulnerabilities are discussed on Twitter at the same time as their public disclosure or shortly afterwards. Moreover, we identified a peak of Tweets in close proximity to the vulnerabilitys public disclosure, which can be traced back to a high number of Re-tweets and tweeting bots (see Figure 3). This is a clear indicator that the information security community tends to discuss and exchange security information during public disclosure more actively.

We found that nearly one quarter of the Tweets follow the *Tweet-Disclosure-Pattern* which means that Tweets discussing certain vulnerabilities appear before their public disclosure or before the availability of countermeasures. According to this observation, security information shared on Twitter can be more current

than conventional, validated information sources, such as vulnerability databases and can serve as potential real-time sensor on insider knowledge about emerging topics in information security. Due to this fact, we empirically confirm the assumption of [34] that vulnerability information appears on Twitter before public disclosure. Consequently, we see a time advantage for security experts to stay informed about emerging vulnerabilities.

As discussed in Section 4.3, 80% of the collected Tweets contain references to websites containing information regarding vulnerabilities, exploits, unofficial countermeasures and official patch releases. According to this observation it can be stated that Twitter serves as crowdsourcing platform were security experts and bots aggregate vulnerability information by referencing a multitude of public available webpages in their Tweets. Consequently, vulnerability information shared on Twitter can be used by organizations to find valuable public available information sources to timely react on newly discovered vulnerabilities and mitigate emerging threats.

A closer look on the different types of referenced information sources (see Section 4.3) showed that links to *demonstrations of exploits* are the most interesting ones. We identified several cases where information on how to exploit a certain vulnerability was posted on Twitter during the black or gray risk phase. For example, we found Tweets referencing Youtube videos demonstrating a certain exploit during black risk phase. According to [14], the point in time during the vulnerability lifecycle where an exploit appears has immense impact on the risk a system affected by certain vulnerabilities is exposed to. A tweet demonstrating an exploit of a vulnerability during the black risk phase is serious since affected organizations might not be aware of the vulnerability and an incident might go unnoticed. In addition, an exploit appearing during gray risk phase might be serious as well, as countermeasures might be not yet available. Moreover, exploit information shared on Twitter before a countermeasure is available are also a serious threat for affected organizations as attackers might use the knowledge to exploit a certain vulnerability. However, exploit information shared on Twitter might be beneficial for organizations as they might be able to stay informed about potential threats and might put suitable countermeasures in place early.

Moreover, our analysis showed that nearly 82% of the Tweets were generated by bots tweeting a reference to a website discussing a specific vulnerability. An analysis of the remaining 18% generated by humans showed a similar pattern. As 80% of all Tweets follow this pattern and the referenced websites by humans and bots appear during all phases of the vulnerability lifecycle and provide valuable information a difference between the human or bot generated contents can not be observed.

5.2 Limitations

Limitations that have to be acknowledged and accounted for regarding our research are: (1) selection bias of relevant Tweets, (2) Twitters data access limit, (3) vague definition of public disclosure, and (4) researchers biasing the analysis of websites through classification mistakes.

In order to counteract (1), we collected all Tweets that contained a valid CVE identifier and cross-validated them with the CVEs contained in the National Vulnerability Database. Tweets that do not contain a CVE identifier were not collected. It is worth mentioning that we primarily focus on CVE-based Tweets which exclude all cyber security-related Tweets that do not reference a respective CVE identifier. There might be the possibility of type (2) limitations, since Twitter limits the public streaming API to only one percent of the daily total number of new Tweets. As the number of daily collected Tweets was below this limit, we were able to crawl all Tweets which were relevant. To overcome (3), we decided to rely on the original definition of public disclosure by [14]. Therefore, we considered the official release data of CVEs on the National Vulnerability Database as public disclosure of a vulnerability to a wider audience. There might be the possibility that researchers made classification mistakes during analysis (cf (4)). We manually classified the 500 most referenced websites. As described in Section 3 two researchers independently classified the websites and the results were compared. If classification discrepancies were discovered they were limited through reclassification.

6 Conclusion and Future Work

In this paper, we present an empirical analysis of the obtainable time advantage of vulnerability information shared on Twitter. We collected 709,880 Tweets between May 23, 2016 and March 27, 2018 and mapped the obtained Tweets to the vulnerability lifecycle model. We observed that one quarter of the examined vulnerabilities were discussed on Twitter before public disclosure by official entities or vendors. Consequently, Twitter can provide a time advantage to react on newly discovered vulnerabilities. Moreover, we observed that Twitter serves as a security crowdsourcing platform for security information which reaches a considerable number of users and organizations. Our analysis identified the following types of information which are referenced in the collected Tweets: (1) Description of vulnerabilities, (2) Demonstrations of exploits, (3) Unofficial proposals of countermeasures, (4) Announcements of patch releases. Future work will focus on social graph analysis of the obtained information in order to identify patterns of collaboration and the development of a prediction model for the severity of vulnerabilities based on the Twitter history of the Tweets authors.

References

1. Al-garadi, M.A., Varathan, K.D., Ravana, S.D.: Cybercrime detection in online communications: The experimental case of cyberbullying detection in the twitter network. Computers in Human Behavior **63**, 433–443 (2016)
2. Arbaugh, W.A., Fithen, W.L., McHugh, J.: Windows of vulnerability: A case study analysis. Computer **33**(12), 52–59 (2000)
3. Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V.: Detecting spammers on twitter. In: Collaboration, electronic messaging, anti-abuse and spam conference (CEAS). vol. 6, p. 12 (2010)

4. Bilge, L., Dumitras, T.: Before we knew it: an empirical study of zero-day attacks in the real world. In: Proceedings of the 2012 ACM conference on Computer and communications security. pp. 833–844. ACM (2012)
5. Bollen, J., Mao, H.: Twitter mood as a stock market predictor. Computer **44**(10), 91–94 (oct 2011). https://doi.org/10.1109/mc.2011.323
6. Cao, Y., Qian, Z., Wang, Z., Dao, T., Krishnamurthy, S.V., Marvel, L.M.: Off-path tcp exploits: Global rate limit considered dangerous. In: 25th USENIX Security Symposium (USENIX Security 16). pp. 210–225 (2016)
7. Chen, C., Wang, Y., Zhang, J., Xiang, Y., Zhou, W., Min, G.: Statistical features-based real-time detection of drifted twitter spam. IEEE Transactions on Information Forensics and Security **12**(4), 914–925 (apr 2017). https://doi.org/10.1109/tifs.2016.2621888
8. Chen, C., Zhang, J., Xiang, Y., Zhou, W., Oliver, J.: Spammers are becoming "smarter" on twitter. IT Professional **18**(2), 66–70 (mar 2016). https://doi.org/10.1109/mitp.2016.36
9. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: #earthquake: Twitter as a distributed sensor system. Transactions in GIS **17**(1), 124–147 (oct 2012)
10. Cui, B., Moskal, S., Du, H., Yang, S.J.: Who shall we follow in twitter for cyber vulnerability? In: Social Computing, Behavioral-Cultural Modeling and Prediction. pp. 394–402. Springer Berlin Heidelberg (2013)
11. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: Botornot: A system to evaluate social bots. In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 273–274. International World Wide Web Conferences Steering Committee (2016). https://doi.org/10.1145/2872518.2889302
12. Erkal, Y., Sezgin, M., Gunduz, S.: A new cyber security alert system for twitter. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE (dec 2015). https://doi.org/10.1109/icmla.2015.133
13. Frei, S., May, M., Fiedler, U., Plattner, B.: Large-scale vulnerability analysis. In: Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense. ACM Press (2006). https://doi.org/10.1145/1162666.1162671
14. Frei, S., Tellenbach, B., Plattner, B.: 0-day patch-exposing vendors(in) security performance. BlackHat Europe (2008)
15. Grier, C., Thomas, K., Paxson, V., Zhang, M.: @ spam: the underground on 140 characters or less. In: Proceedings of the 17th ACM conference on Computer and communications security. pp. 27–37. ACM (2010)
16. Jang-Jaccard, J., Nepal, S.: A survey of emerging threats in cybersecurity. Journal of Computer and System Sciences **80**(5), 973–993 (aug 2014). https://doi.org/10.1016/j.jcss.2014.02.005
17. Jeske, D., McNeill, A.R., Coventry, L., Briggs, P.: Security information sharing via twitter:'heartbleed'as a case study. International Journal of Web Based Communities **13**(2), 172–192 (2017)
18. Khandpur, R.P., Ji, T., Jan, S., Wang, G., Lu, C.T., Ramakrishnan, N.: Crowdsourcing cybersecurity. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management - CIKM 17. ACM Press (2017). https://doi.org/10.1145/3132847.3132866
19. Lee, S., Kim, J.: WarningBird: A near real-time detection system for suspicious URLs in twitter stream. IEEE Transactions on Dependable and Secure Computing **10**(3), 183–195 (may 2013). https://doi.org/10.1109/tdsc.2013.3
20. Mell, P., Grance, T.: Use of the common vulnerabilities and exposures (CVE) vulnerability naming scheme. Tech. rep. (2002). https://doi.org/10.6028/nist.sp.800-51

21. Mell, P.M., Bergeron, T., Henning, D.: Creating a patch and vulnerability management program. Tech. rep. (2005). https://doi.org/10.6028/nist.sp.800-40ver2
22. Mohurle, S., Patil, M.: A brief study of wannacry threat: Ransomware attack 2017. International Journal **8**(5) (2017)
23. Murugan, N.S., Devi, G.U.: Detecting streaming of twitter spam using hybrid method. Wireless Personal Communications (feb 2018). https://doi.org/10.1007/s11277-018-5513-z
24. Ritter, A., Wright, E., Casey, W., Mitchell, T.: Weakly supervised extraction of computer security events from twitter. In: Proceedings of the 24th International Conference on World Wide Web - WWW 15. ACM Press (2015). https://doi.org/10.1145/2736277.2741083
25. Sabottke, C., Suciu, O., Dumitra, T.: Vulnerability disclosure in the age of social media: exploiting twitter for predicting real-world exploits. In: 24th USENIX Security Symposium (USENIX Security 15). pp. 1041–1056 (2015)
26. Sauerwein, C., Sillaber, C., Breu, R.: Shadow cyber threat intelligence and its use in information security and risk management processes. In: Multikonferenz Wirtschaftsinformatik (MKWI 2018) (2018)
27. Sauerwein, C., Sillaber, C., Mussmann, A., Breu, R.: Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. Proceedings of the International Conference on Wirtschaftsinformatik 2017 (WI 2017) (2017)
28. Sceller, Q.L., Karbab, E.B., Debbabi, M., Iqbal, F.: SONAR. In: Proceedings of the 12th International Conference on Availability, Reliability and Security. ACM Press (2017)
29. Shen, H., Liu, X.: Detecting spammers on twitter based on content and social interaction. In: 2015 International Conference on Network and Information Systems for Computers. IEEE (jan 2015). https://doi.org/10.1109/icnisc.2015.82
30. Soomro, Z.A., Shah, M.H., Ahmed, J.: Information security management needs more holistic approach: A literature review. International Journal of Information Management **36**(2), 215–225 (apr 2016). https://doi.org/10.1016/j.ijinfomgt.2015.11.009
31. Stilo, G., Velardi, P., Tozzi, A.E., Gesualdo, F.: Predicting flu epidemics using twitter and historical data. In: Brain Informatics and Health. pp. 164–177. Springer International Publishing (2014)
32. Suárez-Serrato, P., Roberts, M.E., Davis, C., Menczer, F.: On the influence of social bots in online protests. In: Lecture Notes in Computer Science. pp. 269–278. Springer International Publishing (2016)
33. Syed, R.: Analyzing software vendors patch release behavior in the age of social media. Proceedings of the International Conference on Information Systems 2017 (ICIS 2017) (2017)
34. Syed, R., Rahafrooz, M., Keisler, J.M.: What it takes to get retweeted: An analysis of software vulnerability messages. Computers in Human Behavior **80**, 207–215 (mar 2018). https://doi.org/10.1016/j.chb.2017.11.024
35. Trabelsi, S., Plate, H., Abida, A., Aoun, M.M.B., Zouaoui, A., Missaoui, C., Gharbi, S., Ayari, A.: Mining social networks for software vulnerabilities monitoring. In: 2015 7th International Conference on New Technologies, Mobility and Security (NTMS). IEEE (jul 2015). https://doi.org/10.1109/ntms.2015.7266506
36. Wang, A.H.: Don't follow me: Spam detection in twitter. In: Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on. pp. 1–10. IEEE (2010). https://doi.org/10.5220/0002996201420151