International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

AUTOMATED IDENTIFICATION AND PROFILING OF EMERGING CYBER THREATS USING NATURAL LANGUAGE PROCESSING

[1]M SRIDHAR KUMAR, [2]MANGALARAPU SRIVANI
[1]Assistant Professor,[2]Student
Department of CSE
Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT

The amount of time that elapses between the discovery of a new cyber vulnerability and its exploitation by malevolent actors online has been decreasing over time. A recent example that effectively shows this notion is the Log4j vulnerability. Hackers quickly launched network-wide scans to find weak hosts in order to install malicious software, such as ransomware and bitcoin miners, on sensitive devices as soon as the vulnerability was made public. Consequently, in order to maximise the efficacy of preventative measures, the cybersecurity defence plan must quickly detect threats and their capabilities. For security analysts, spotting emerging risks is a crucial responsibility, but it may be challenging as it requires them to sift through a lot of data and information from many sources. Here, we provide a system that uses Twitter tweets as an event source and MITRE ATT&CK as a knowledge source to automatically identify and characterise new hazards. The three main parts of the system are: identifying and labelling cyberthreats; using two machine learning layers to filter and classify tweets in order to profile the identified threat in terms of its goals or objectives; and raising alerts based on the threat's level of risk. Our research's main contribution is the way we analysed and characterised the threats we found by classifying them according to their goals or intentions. This method provides more understanding about the nature of the threat and makes recommendations for potential countermeasures. In our research, the profiling step successfully identified and categorised risks with a 77% F1 score.

1. INTRODUCTION

A tendency towards hyper-connectivity and hyper-mobility has led to a recent increase in the dependence on the Internet for social interactions, commerce, and governance. Although the Internet has developed into a vital infrastructure for organisations, governments, and society, there is also a greater chance of cyberattacks with various goals and objectives. It takes timely knowledge about cyber vulnerabilities and assaults, sometimes known as threats, to protect organisations against cyber exploitation [1].

The definition of threat intelligence is described as "evidence-based knowledge

Index in Cosmos
Oct 2024, Volume 14, ISSUE 4
UGC Approved Journal

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard, including context, mechanisms, indicators, implications, and actionable advice." [2]. Cyber threat intelligence, also known as threat intelligence in the cyber security realm, offers pertinent and timely information, such as attack signatures, that may assist lower the uncertainty associated with spotting possible security flaws and assaults.

Informal or governmental sources that formally disseminate threat information in structured data format are often good places to get cyber threat intelligence. Structured threat intelligence follows a standard format and data model that is well-defined. Therefore, security systems can simply parse structured cyber threat information to analyse and appropriately react to security risks. The Common Vulnerabilities and Exposures (CVE) database1 and the National Vulnerability Database (NVD) are two official sources of cyber threat information.2.

Informal sources such as social media platforms, forums, dark webs, and open blogs may also provide cyber threat information. Informal sources enable any individual or organisation to instantly disseminate threat intelligence in unstructured data format or natural language on the Internet. Open Source Intelligence (OSINT) is another term for threat intelligence that is unstructured and accessible to the general public [3]. Early warning systems for cyber security events, such as security vulnerability exploits, are provided by cyber security-related OSINT [4].

Malicious actors usually need to do the following in order to carry out a cyber-attack: 1) locate vulnerabilities; 2) get the tools and know-how needed to exploit them; 3) choose a target and enlist allies; 4) build or buy the infrastructure required; and 5) plan and carry out the assault. System administrators, security analysts, and even victims are examples of other players who could talk about vulnerabilities or plan an attack response [5]. Digital traces are left behind by these actions, which are often carried out online via social media, professional blogs, and open and closed Web forums. When taken as a whole, these digital footprints provide insightful information about how cyber risks are changing and may warn of an impending or ongoing assault long before any malicious behaviour is discovered on the target machine. For example, vulnerabilities are discussed on dark web forums even before they are discussed on social media [6] and on Twitter before they are made public [4].

2. EXISTING SYSTEM

Cybersecurity is becoming an ever-increasing concern for most organizations and much research has been developed in this field over the last few years. Inside these organizations, the Security Operations

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Center (SOC) is the central nervous system that provides the necessary security against cyber threats. However, to be effective, the SOC requires timely and relevant threat intelligence to accurately and properly monitor, maintain, and secure an IT infrastructure. This leads security analysts to strive for threat awareness by collecting and reading various information feeds. However, if done manually, this results in a tedious and extensive task that may result in little knowledge being obtained given the large amounts of irrelevant information. Research has shown that Open Source Intelligence (OSINT) provides useful information to identify emerging cyber threats.

OSINT is the collection, analysis, and use of data from openly available sources for intelligence purposes [21]. Examples of sources for OSINT are public blogs, dark and deep websites, forums, and social media. In such platforms, any person or entity on the Internet can publish, in real-time, information in natural language related to cyber security, including incidents, new threats, and vulnerabilities. Among the OSINT sources for cyber threat intelligence, we can highlight the social media Twitter as one of the most representative [22]. Cyber security experts, system administrators, and hackers constantly use Twitter to discuss technical details about cyberattacks and share their experiences [4].

Utilization of OSINT to automatically identify cyber threats via social media, forums and other openly available sources using text analytics was proposed in different researches [1], [23], [7], [24], [25], [26], [13], [27] and [28]. However, most proposals focus on identifying important events related to cyber threats or vulnerabilities but do not propose identifying and profiling cyber threats.

Amongst research, [13] proposes an early cyber threat warning system that mines online chatter from cyber actors on social media, security blogs, and dark web forums to identify words that signal potential cyber-attacks. The framework is comprised by two main components: text mining and warning generation. The text mining phase consists on pre-processing the input data to identify potential threat names by discarding ‘‘known’’ terms and selecting repeating ‘‘unknown’’ among different sources as they potentially can be the name of a new or discovered cyber threat. The second component, warning generation, is responsible for issuing alarms for unknown terms that meet some requirements, like appearing twice in a given period of time. The approach presented in this research uses keyword filtering as the only strategy to identify cyber threat names, which may result in false positives as unknown words may appear in tweets or other content not necessarily related to cyber security. Additionally, this research does not profile the identified cyber threat.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

First, the proposed approach does not name the identified threat. Naming the threat is an important step to cyber threat intelligence
as it may allow analysts to identify and mitigate campaigns based on the historic modus operandi employed by a given threat or group.

Second, the proposed approach relies on an external component to classify tweets as related or not to cyber security as opposed to our approach that proposes a component to classify tweets using machine learning trained with the evolving knowledge from MITRE ATT&CK. Third, instead of using a keyword match to pre-filter threats and a fixed list of threat types, we present an approach to profile the identified cyber threat to spot in which phase of phases of the cyber kill chain the given threat operates in. This is important for a cyber threat analyst as he or she may employ the necessary mitigation steps depending on the threat profile.

In [1], a framework for automatically gathering cyber threat intelligence from Twitter is presented. The framework utilizes a novelty detection model to classify the tweets as relevant or irrelevant to Cyber threat intelligence. The novelty classifier learns the features of cyber threat intelligence from the threat descriptions in the Common Vulnerabilities and Exposures (CVE) database 5 and classifies a new unseen tweet as normal or abnormal in relation to Cyber threat intelligence. The normal tweets are considered as Cyber threat relevant while the abnormal tweets are considered as Cyber threat-irrelevant. The paper evaluates the framework on a data set created from the tweets collected over a period of twelve months in 2018 from 50 influential Cyber security-related accounts. During the evaluation, the framework achieved the highest performance of 0.643 measured by the F1-score metric for classifying Cyber threat tweets. According to the authors, the proposed approach outperformed several
baselines including binary classification models. Also, was analyzed the correctly classified cyber threat tweets and discovered that 81 of them do not contain a CVE identifier. The authors have also found that 34 out of the 81 tweets can be associated with a CVE identifier included in the top 10 most similar CVE descriptions of each tweet. Despite presenting a proposal to distinguish between relevant and irrelevant tweets, the proposal does not address the identification of threats and their intentions. Those are important requirements for Cyber Threat Intelligence in formulating defense strategies against emerging threats.

The tool proposed in [23] collects tweets from a selected subset of accounts using the Twitter streaming API, and then, by using keyword-based filtering, it discards tweets unrelated to the monitored infrastructure assets. To classify and extract information from tweets the paper uses a sequence of

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

two deep neural networks. The first is a binary classifier based on a Convolutional Neural Network (CNN) architecture used for Natural Language Processing (NLP) [29]. It receives tweets that may be referencing an asset from the monitored infrastructure and labels them as either relevant when the tweets contain security-related information, or irrelevant otherwise.

Disadvantages

- An existing system never implemented Multi-Class machine learning (ML) algorithms - the next steps in the pipeline.
- An existing system didn't implement the following method PROCESS IDENTIFIED AND CLASSIFIED THREATS.

3. PROPOSED SYSTEM

The overall goal of this work is to propose an approach to automatically identify and profile emerging cyber threats based on OSINT (Open Source Intelligence) in order to generate timely alerts to cyber security engineers. To achieve this goal, we propose a solution whose macro steps are listed below.

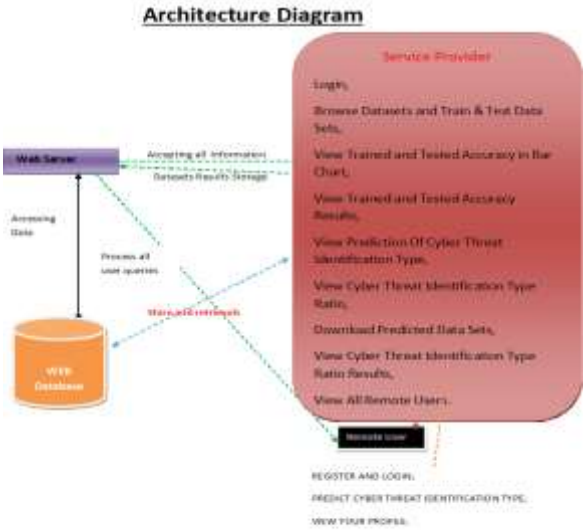
1) Continuously monitoring and collecting posts from prominent people and companies on Twitter to mine unknown terms related to cyber threats and malicious campaigns;
2) Using Natural Language Processing (NLP) and Machine Learning (ML) to identify those terms most likely to be threat names and discard those least likely;
3) Leveraging MITRE ATT&CK techniques' procedures examples to identify most likely tactic employed by the discovered threat;
4) Generating timely alerts for new or developing threats along with its characterization or goals associated with a risk rate based on how fast the threat is evolving since its identification.

Advantages

To conduct a cyber-attack, malicious actors typically have to
1) Identify vulnerabilities,
2) acquire the necessary tools and tradecraft to successfully exploit them,
3) choose a target and recruit participants,
4) Create or purchase the infrastructure needed, and
5) Plan and execute the attack. Other actors— system administrators, security analysts, and even victims— may discuss vulnerabilities or coordinate a response to attacks

SYSTEM ARCHITECTURE

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86



4. ALGORITHIM

Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

Logistic regression Classifiers

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an independent and identically distributed (iid) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to genetic algorithms (GAs) or perceptrons, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a type of deep learning algorithm specifically designed for image processing and recognition tasks. Compared to alternative classification models, CNNs require less preprocessing as they can automatically learn hierarchical feature representations from raw input images. They excel at assigning importance to various objects and features within the images through convolutional layers, which apply filters to detect local patterns.

The connectivity pattern in CNNs is inspired by the visual cortex in the human brain, where neurons respond to specific regions or receptive fields in the visual space. This architecture enables CNNs to effectively capture spatial relationships and patterns in images. By stacking multiple convolutional

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

and pooling layers, CNNs can learn increasingly complex features, leading to high accuracy in tasks like image classification, object detection, and segmentation.

5. IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Cyber Threat Identification Type, View Cyber Threat Identification Type Ratio, Download Predicted Data Sets, View Cyber Threat Identification Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT CYBER THREAT IDENTIFICATION TYPE, VIEW YOUR PROFILE.

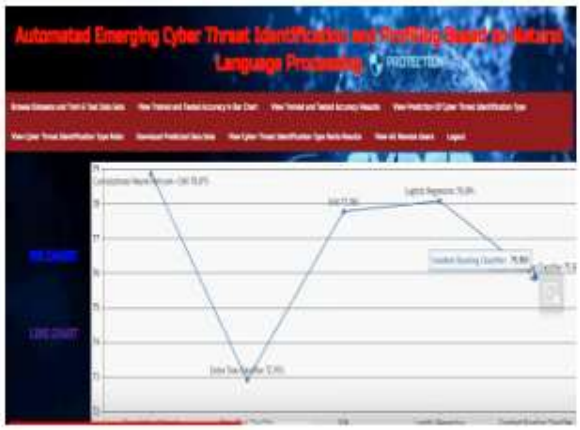
6. SCREEN SHOTS

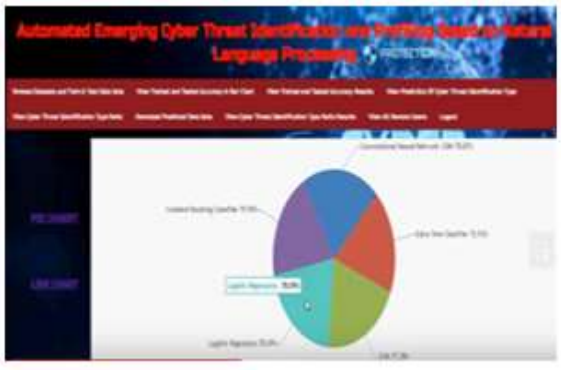
Index in Cosmos
Oct 2024, Volume 14, ISSUE 4
UGC Approved Journal

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Bar Graph: Represent the distribution of identified threats by type or category, aiding in understanding the prevalence of different threat categories.

Line Charts Graph: Depict the trend of emerging threats over time, showcasing the frequency or intensity of threats detected.





7. CONCLUSION

Keeping abreast of emerging vulnerabilities and threats is a difficult but crucial duty for analysts, given the dynamic nature of the cyber security industry. Even with the strongest processes and controls in place, a novel danger can emerge and find a novel way around the defences, necessitating an immediate reaction. In this sense, up-to-date knowledge of newly developing cyberthreats

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

becomes critical to a functioning cyber security system.

This study suggests an automated method for identifying and characterising cyber threats by analysing Twitter conversations using natural language processing. The goal is to precisely collaborate with the diligent effort of monitoring Twitter, a wealth of information, in order to promptly extract important information about new dangers.

This work sets itself apart from others by not stopping at just pointing out the danger. By comparing the language from tweets to the actions taken by actual threats as detailed in the MITRE ATT&CK knowledge base, it attempts to determine the threat's objectives. By using this dynamic and cooperative knowledge base to train machine learning algorithms, the cyber security community's efforts may be harnessed to automatically profile detected cyber threats according to their intentions.

In addition to conducting the research experiment, we put our strategy to the test by implementing the suggested pipeline and running it for 70 days, producing online warnings for the Threat Intelligence Team of a significant Brazilian financial institution. At least three threats during this time prompted the team to take precautionary measures; one such instance is the Petit Potam case, which is covered in section V. Petit-Potam was brought to the team's attention by our system 17 days before to Microsoft's official patch release. The defence team was able to put mitigations in place within this time frame, preventing possible exploits and events as a result.

Our tests revealed that the profiling stage achieved an F1 score of 77% in accurately identifying threats among 14 distinct approaches, with a 15% false alarm rate. Future work should focus on improving the false positives rate in the tweet selection phases (Unknown Words and One-class) and increasing the accuracy of the approach linked with the detected threat in the profiling stage. We are attempting to address this by using the part of speech (POS) algorithm implementation from the Spacy29 Python library in an experiment with an alternative NLP methodology. The goal is to choose tweets where the activity described (the root verb) refers to the unknown term by identifying the root verb, the subject, and the object of the sentences.

REFERENCES

[1] B. D. Le, G. Wang, M. Nasim, and A. Babar, ‘‘Gathering cyber threat intelligence from Twitter using novelty classification,’’ 2019, *arXiv:1907.01755*.

[2] *Definition: Threat Intelligence*, Gartner Research, Stamford, CO, USA, 2013.

[3] R. D. Steele, ‘‘Open source intelligence: What is it? why is it important to the military,’’ *Journal*, vol. 17, no. 1, pp. 35–41, 1996.

[4] C. Sabottke, O. Suciu, and T. Dumitras, ‘‘Vulnerability disclosure in the age of social media: Exploiting Twitter for

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

predicting real-world exploits,'' in *Proc. 24th USENIX Secur. Symp. (USENIX Secur.)*, 2015, pp. 1041–1056.

[5] A. Sapienza, A. Bessi, S. Damodaran, P. Shakarian, K. Lerman, and E. Ferrara, ''Early warnings of cyber threats in online discussions,'' in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 667–674.

[6] E. Nunes, A. Diab, A. Gunn, E. Marin, V. Mishra, V. Paliath, J. Robertson, J. Shakarian, A. Thart, and P. Shakarian, ''Darknet and deepnet mining for proactive cybersecurity threat intelligence,'' in *Proc. IEEE Conf. Intell. Secur. Informat. (ISI)*, Sep. 2016, pp. 7–12.

[7] S. Mittal, P. K. Das, V. Mulwad, A. Joshi, and T. Finin, ''CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities,'' in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 860–867.

[8] A. Attarwala, S. Dimitrov, and A. Obeidi, ''How efficient is Twitter: Predicting 2012 U.S. presidential elections using support vector machine via Twitter and comparing against Iowa electronic markets,'' in *Proc. Intell. Syst. Conf. (IntelliSys)*, Sep. 2017, pp. 646–652.

[9] N. Dionísio, F. Alves, P. M. Ferreira, and A. Bessani, ''Towards end-to-end cyberthreat detection from Twitter using multi-task learning,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8. [10] O. Oh, M. Agrawal, and H. R. Rao, ''Information control and terrorism: Tracking the Mumbai terrorist attack through Twitter,'' *Inf. Syst. Frontiers*, vol. 13, no. 1, pp. 33–43, Mar. 2011.

[11] T. Sakaki, M. Okazaki, and Y. Matsuo, ''Earthquake shakes Twitter users: Real-time event detection by social sensors,'' in *Proc. 19th Int. Conf.World Wide Web*, Apr. 2010, pp. 851–860.

[12] B. De Longueville, R. S. Smith, and G. Luraschi, '''OMG, from here, I can see the flames!': A use case of mining location based social networks to acquire spatio-temporal data on forest fires,'' in *Proc. Int. Workshop Location Based Social Netw.*, Nov. 2009, pp. 73–80.

[13] A. Sapienza, S. K. Ernala, A. Bessi, K. Lerman, and E. Ferrara, ''DISCOVER: Mining online chatter for emerging cyber threats,'' in *Proc. Companion Web Conf. Web Conf. (WWW)*, 2018, pp. 983–990.

[14] R. P. Khandpur, T. Ji, S. Jan, G. Wang, C.-T. Lu, and N. Ramakrishnan, ''Crowdsourcing cybersecurity: Cyber attack detection using social media,'' in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 1049–1057.

[15] Q. Le Sceller, E. B. Karbab, M. Debbabi, and F. Iqbal, ''SONAR: Automatic detection of cyber security events over the Twitter stream,'' in *Proc. 12th Int. Conf. Availability, Rel. Secur.*, Aug. 2017, pp. 1–11.

[16] K.-C. Lee, C.-H. Hsieh, L.-J.Wei, C.-H. Mao, J.-H. Dai, and Y.-T. Kuang, ‘‘Sec-buzzer: Cyber security emerging topic mining with open threat intelligence retrieval and timeline event annotation,’’ *Soft Comput.*, vol. 21, no. 11, pp. 2883–2896, Jun. 2017.

[17] A. Ritter, E. Wright, W. Casey, and T. Mitchell, ‘‘Weakly supervised extraction of computer security events from Twitter,’’ in *Proc. 24th Int. Conf. World Wide Web*, May 2015, pp. 896–905.

[18] A. Queiroz, B. Keegan, and F. Mtenzi, ‘‘Predicting software vulnerability using security discussion in social media,’’ in *Proc. Eur. Conf. Cyber Warfare Secur.*, 2017, pp. 628–634.

[19] A. Bose, V. Behzadan, C. Aguirre, and W. H. Hsu, ‘‘A novel approach for detection and ranking of trendy and emerging cyber threat events in Twitter streams,’’ in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2019, pp. 871–878.

[20] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas, ‘‘Mitre ATT&CK: Design and philosophy,’’ MITRE Corp., McLean, VA, USA, Tech. Rep. 19-01075-28, 2018.