EMERGING CYBER THREAT DETECTION AND PROFILING THROUGH AUTOMATED NATURAL LANGUAGE PROCESSING

[1]Dr O. Sampath, [2]Biyyala Aravind
[1]Associate Professor, [2]MCA Student
Department of Master of Computer Application, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, 518501, Andhra Pradesh, India.

ABSTRACT
The temporal gap between the revelation of a novel cyber vulnerability and its exploitation by cyber malefactors has been progressively diminishing throughout the course of time. The Log4j vulnerability is a recent example that illustrates this point well. Shortly after the vulnerability was made public, hackers promptly initiated internet-wide scans to identify susceptible hosts for installing malicious software like as bitcoin miners and ransomware on vulnerable devices. Therefore, it is crucial for the cybersecurity defense strategy to promptly identify threats and their capabilities in order to optimize the effectiveness of preventive measures. Identifying new threats is a critical task for security analysts, but it is difficult since they have to analyze a large amount of data and information from many sources to detect any developing threats. Here, we provide a system that automatically identifies and profiles emerging risks by use Twitter posts as a source of events and MITRE ATT&CK as a source of knowledge for characterizing threats. The system consists of three primary components: detection of cyber threats and their names; profiling the discovered threat in terms of its objectives or aims using two machine learning layers to filter and categorize tweets; and generating alarms depending on the danger associated with the threat. The primary contribution of our research is the method used to analyze and describe the discovered threats based on their intents or objectives. This approach offers more insight into the nature of the danger and suggests possible ways to address it. During our studies, the profiling stage achieved an F1 score of 77% in accurately identifying and categorizing detected threats.

1. INTRODUCTION

Recently there has been an increasing reliance on the Internet for business, government, and social interactions as a result of a trend of hyper-connectivity and hyper-mobility. While the Internet has become an indispensable infrastructure for businesses, governments, and societies, there is also an increased risk of cyberattacks with different motivations and intentions. Preventing organizations from cyber exploits needs timely intelligence about cyber vulnerabilities and attacks, referred to as threats [1].

Threat intelligence is defined as ‘‘evidence-based knowledge, including context, mechanisms, indicators, implications, and actionable advice, 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'' [2]. Threat intelligence in cyber security domain, or cyber threat intelligence, provides timely and relevant information, such as signatures of the attacks, that can help reduce the uncertainty in identifying potential security vulnerabilities and attacks.

Cyber threat intelligence can generally be extracted from informal or formal sources, which officially release threat information in structured data format. Structured threat intelligence adheres to a well-defined data model, with a common format and structure. Structured cyber threat intelligence, therefore, can be easily parsed by security tools to analyze and respond to security threats accordingly. Examples of formal sources of cyber threat intelligence include the Common Vulnerabilities and Exposures (CVE) database1 and the National Vulnerability Database (NVD).2

Cyber threat intelligence is also available on informal sources, such as public blogs, dark webs, forums, and social media platforms. Informal sources allow any person or entity on the Internet to publish, in real-time, the threat information in natural language, or unstructured data format. The unstructured and publicly available threat intelligence is also called Open-Source Intelligence (OSINT) [3]. Cyber security-related OSINT are early warning sources for cyber security events such as security vulnerability exploits [4].

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 [5]. These activities are often conducted online through social media, (open and dark) Web forums, and professional blogs, leaving digital traces behind. Collectively, these digital traces provide valuable insights into evolving cyber threats and can signal a pending or developing attack well before the malicious activity is noted on a target system. For example, exploits are discussed on Twitter before they are publicly disclosed and on dark web forums even before they are discussed on social media [6].

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 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 [7]. 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 [8]. Cyber security experts, system administrators, and hackers constantly use Twitter to discuss technical details about cyberattacks and share their experiences [9].

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], [7], [13], [23], [24], [25], [26], [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, [10] 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.

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 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.

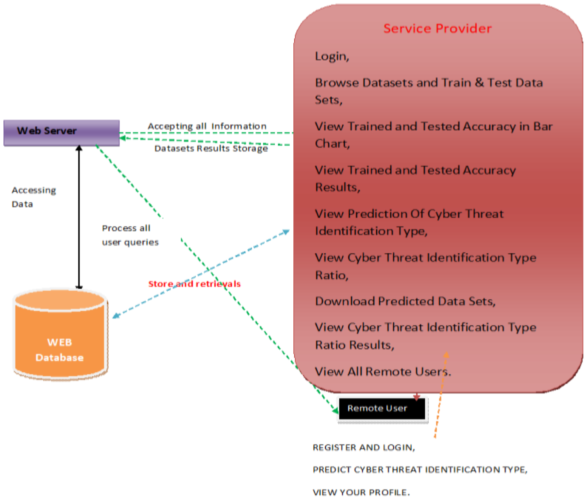
3. PROPOSED SYSTEM

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.

SYSTEM ARCHITECTURE

Architecture Diagram


Service Provider
Login,
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.
Web Server
Accepting all Information
Datasets Results Storage
Accessing Data
Process all user queries
Store and retrievals
WEB Database
Remote User
REGISTER AND LOGIN,
PREDICT CYBER THREAT IDENTIFICATION TYPE,
VIEW YOUR PROFILE.

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.

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 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 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 perceptron's, both of which are widely used for classification in machine learning. For perceptron's, 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 perceptron's 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 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.

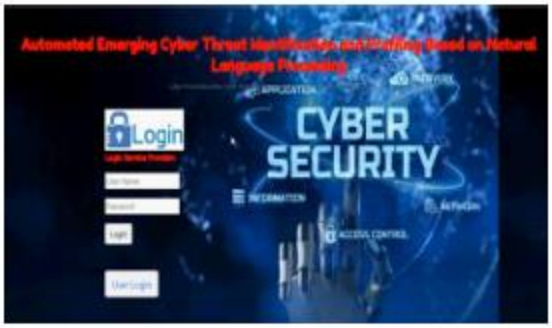
View and Authorize Users

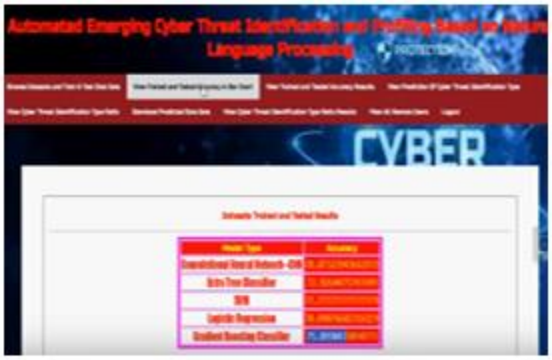
In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorize the users.

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.

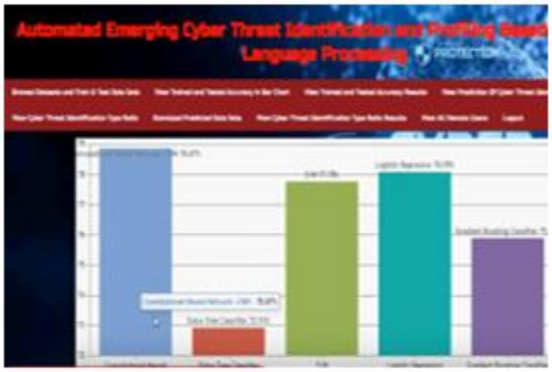
6. SCREEN SHOTS


Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
Login
CYBER
SECURITY
APPLICATION
INFORMATION
ACCESS CONTROL
Login
User Login

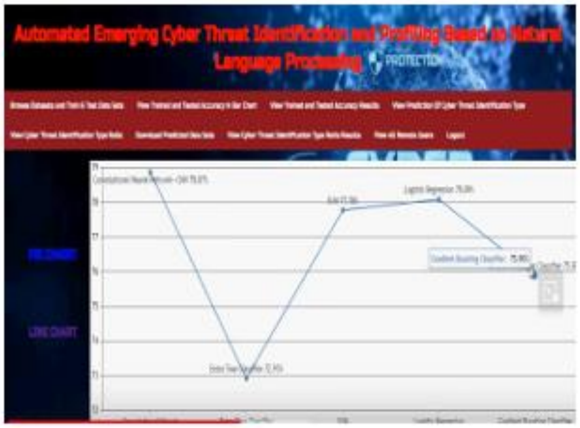

Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION
CYBER

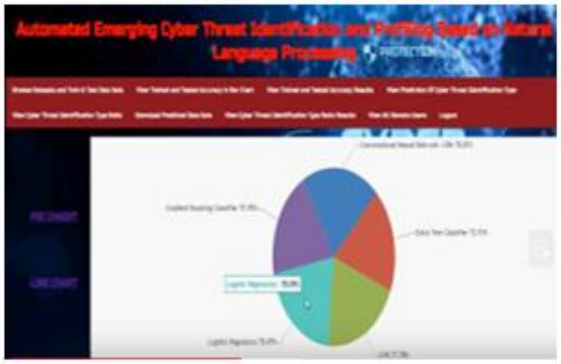
Datasets Trained and Tested Results

Model Type	Accuracy
Convolutional Neural Network - CNN	[illegible]
Extra Tree Classifier	[illegible]
SVM	[illegible]
Logistic Regression	[illegible]
Gradient Boosting Classifier	[illegible]


Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION

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


Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION
PIE CHART
LINE CHART


Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION
PIE CHART
LINE CHART

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

Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION
CYBER

View Cyber Threat Identification Type Ratio Details

Cyber Threat Identification Type	Ratio
Cyber Threat Found	[illegible]
No Cyber Threat Found	[illegible]

Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION
CYBER

View Cyber Threat Identification Type Ratio Details

Cyber Threat Identification Type	Ratio
Cyber Threat Found	[illegible]
No Cyber Threat Found	[illegible]

7. CONCLUSION

Given the dynamism of the cyber security field, with new vulnerabilities and threats appearing at any time, keeping up to date on them is a challenging but important task for analysts. Even following the best practices and applying the best controls, a new threat may bring an unusual way to subvert the defenses requiring a quick response. This way, timely information about emerging cyber threats becomes paramount to a complete cyber security system.

This research proposes an automated cyber threat identification and profiling based on the natural language processing of Twitter messages. The objective is exactly to cooperate with the hard work of following the rich source of information that is Twitter to extract valuable information about emerging threats in a timely manner.

This work differentiates itself from others by going a step beyond identifying the threat. It seeks to identify the goals of the threat by mapping the text from tweets to the procedures conducted by real threats described in MITRE ATT&CK knowledge base. Taking advantage of this evolving and collaborative knowledge base to train machine learning algorithms is a way to leverage the efforts of cyber security community to automatically profile identified cyber threats in terms of their intents.

To put in test our approach, in addition to the research experiment, we implemented the proposed pipeline and run it for 70 days generating online alerts for the Threat Intelligence Team of a big financial institution in Brazil. During this period, at least three threats made the team take preventive actions, such as the Petit Potam case, described in section V. Our system alerted the team making them aware of Petit-Potam 17 days before the official patch was published by Microsoft. Within this period, the defense team was able to implement mitigations avoiding potential exploits and, consequently, incidents.

Our experiments showed that the profiling stage reached an F1 score of 77% in correctly profiling discovered threats among 14 different tactics and the percentage of false alerts of 15%. In future work, we consider it important to advance in tweets selection stages (Unknown Words and One-class), to improve the false positives rate and in the profiling stage, to reach higher accuracy in determining the technique associated with the identified threat. We are working on this way by experimenting with a different NLP approach using the part of speech (POS) algorithm implementation from Spacy29 Python library. The object is to identify the root verb, the subject, and the object of the phrases to select tweets where the action described (the root verb) is referencing the unknown word.

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 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. Bess ani, ''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.