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

[1] J.Rakesh Babu, Assistant Professor, Department of CSE, Chalapathi Institute of Technology, Guntur.
[2] Korlapati Kushitha, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
[3] Alaparthi Sumanth, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
[4] Marri Ithamraju, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
[5] Gumma Siva Shankar, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.

Abstract: The time window between the disclosure of new cyber vulnerability and its use by
Cyber criminals have been getting smaller and smaller over time. Recent episodes, such as Log4j vulnerability, exemplify this well. Within hours after the exploit being released, attackers started scanning the internet looking for vulnerable hosts to deploy threats like crypto currency miners and ransom ware on vulnerable systems. Thus, it becomes imperative for the cyber security defense strategy to detect threats and their capabilities as early as possible to maximize the success of prevention actions. Although crucial, discovering new threat s is a challenging activity for security analysts due to the immense volume of data and information sources to be analyzed for signs that a threat is emerging. In this sense, we present a framework for automatic identification and profiling of emerging threats using Twitter messages as a source of events and MITRE ATT&CK as a source of knowledge for threat characterization. The framework comprises three main parts: identification of cyber threats and their names; profiling the identified threat in terms of its intentions or goals by employing two machine learning layers to filter and classify tweets; and alarm generation based on the threat's risk. The main contribution of our work is the approach to characterize or profile the identified threats in terms of their intentions or goals, providing additional context on the threat and avenues for mitigation. In our experiments, the profiling stage reached an F1 score of 77% in correctly profiling discovered 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 cyber attacks 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). 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 [4] and on dark web forums even before they are discussed on social media [6].

2. LITERATURE SURVEY

1) Artificial intelligence for cyber security: Literature review and future research directions
Abstract: Artificial intelligence (AI) is a powerful technology that helps cyber security teams automate repetitive tasks, accelerate threat detection and response, and improve the accuracy of their actions to strengthen the security posture against various security issues and cyber attacks. This article presents a systematic literature review

and a detailed analysis of AI use cases for cyber security provisioning. The review resulted in 2395 studies, of which 236 were identified as primary. This article classifies the identified AI use cases based on a NIST cyber security framework using a thematic analysis approach. This classification framework will provide readers with a comprehensive overview of the potential of AI to improve cyber security in different contexts. The review also identifies future research opportunities in emerging cyber security application areas, advanced AI methods, data representation, and the development of new infrastructures for the successful adoption of AI-based cyber security in today's era of digital transformation and polycrisis
.

2) NLP-Based Techniques for Cyber Threat Intelligence

Abstract: In the digital era, threat actors employ sophisticated techniques for which, often, digital traces in the form of textual data are available. Cyber Threat Intelligence (CTI) is related to all the solutions inherent to data collection, processing, and analysis useful to understand a threat actor's targets and attack behavior. Currently, CTI is assuming an always more crucial role in identifying and mitigating threats and enabling proactive defense strategies. In this context, NLP, an artificial intelligence branch, has emerged as a powerful tool for enhancing threat intelligence capabilities. This survey paper provides a comprehensive overview of NLP-based techniques applied in the context of threat intelligence. It begins by describing the foundational definitions and principles of CTI as a major tool for safeguarding digital assets. It then undertakes a thorough examination of NLP-based techniques for CTI data crawling from Web sources, CTI data analysis, Relation Extraction from cyber security Data, CTI sharing and collaboration, and security threats of CTI. Finally, the challenges and limitations of NLP in threat intelligence are exhaustively examined, including data quality issues and ethical considerations. This survey draws a complete framework and serves as a valuable resource for security professionals and researchers seeking to understand the state-of-the-art NLP-based threat intelligence techniques and their potential impact on cyber security

3) A Systematic Literature Review on Cyber Threat Intelligence for Organizational Cyber security Resilience

Abstract: Cyber security is a significant concern for businesses worldwide, as cybercriminals target business data and system resources. Cyber threat intelligence (CTI) enhances organizational cyber security resilience by obtaining, processing, evaluating, and disseminating information about potential risks and opportunities inside the cyber domain. This research investigates how companies can employ CTI to improve their precautionary measures against security breaches. The study follows a systematic review methodology, including selecting primary studies based on specific criteria and quality valuation of the selected papers. As a result, a comprehensive framework is proposed for implementing CTI in organizations. The proposed framework is comprised of a knowledge base, detection models, and visualization dashboards. The detection model layer consists of behavior-based, signature-based, and anomaly-based detection. In contrast, the knowledge base layer contains information resources on possible threats, vulnerabilities, and dangers to key assets. The visualization dashboard layer provides an overview of key metrics related to cyber threats, such as an organizational risk meter, the number of attacks detected, types of attacks, and their severity level. This relevant systematic study also provides insight for future studies, such as how organizations can tailor their approach to their needs and resources to facilitate more effective collaboration between stakeholders while navigating legal/regulatory constraints related to information sharing.

4) Explainable artificial intelligence for cybersecurity: a literature survey

Abstract: With the extensive application of deep learning (DL) algorithms in recent years, e.g., for detecting Android malware or vulnerable source code, artificial intelligence (AI) and machine learning (ML) are increasingly becoming essential in the development of cybersecurity solutions. However, sharing the same fundamental limitation with other DL application domains, such as computer vision (CV) and natural language processing (NLP), AI-based cybersecurity solutions are incapable of justifying the results (ranging from detection and prediction to reasoning and decision-making) and making them understandable to humans. Consequently, explainable AI (XAI) has emerged as a paramount topic addressing the related challenges of making AI models explainable or interpretable to human users. It is particularly relevant in cyber security domain, in that XAI may allow security operators, who are overwhelmed with tens of thousands of security alerts per day (most of which are false positives), to better assess the potential threats and reduce alert fatigue. We conduct an extensive literature review on the intersection between XAI and cyber security. Particularly, we investigate the existing literature from two perspectives: the applications of XAI to cyber security (e.g., intrusion detection, malware classification), and the security of XAI (e.g., attacks on XAI pipelines, potential countermeasures). We characterize the security of XAI with several security properties that have been discussed in the literature. We also formulate open questions that are either unanswered or insufficiently addressed in the literature, and discuss future directions of research.

3. EXISTING SYSTEM

Cyber security 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 [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 cyber attacks 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.

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.

4. 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

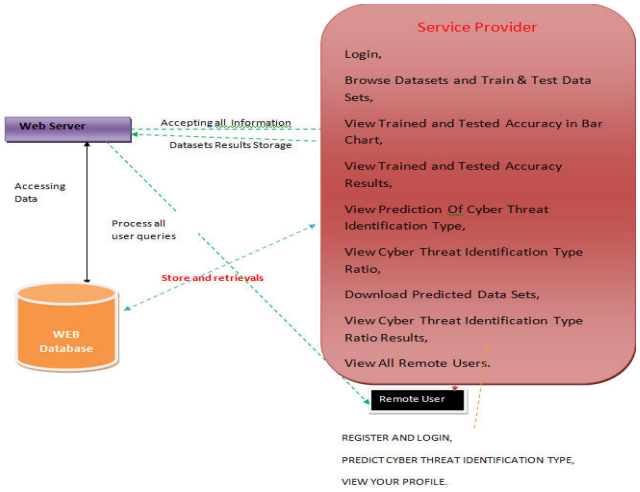

Fig 1: System Architecture

5. ALGORITHMS

5.1 DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

5.2 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 discriminate 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 discriminate analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminate 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.

5.3 SVM

In classification tasks a discriminate machine learning technique aims at finding, based on an independent and identically distributed (iid) training dataset, a discriminate function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminate 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, discriminate 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 discriminate technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyper plane parameter—in contrast to genetic algorithms (GAs) or perceptions, both of which are widely used for classification in machine learning. For perceptions, 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 perception and GA classifier models are different each time training is initialized. The aim of GAs and perceptions is only to minimize error during training, which will translate into several hyper planes' meeting this requirement.

5.4 RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the

random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance. Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

6. RESULTS

6.1 Output Screens

Fig 6.1 Login Page

In above screen is the Login page

Fig 6.2 Accuracy for the ml algorithms

In above screen shows the different machine learning algorithms accuracy.

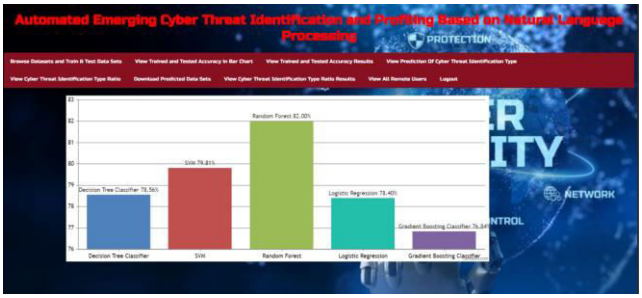

Fig 6.3 Accuracy in Bar Charts

In above screen shows algorithms accuracy in bar charts.

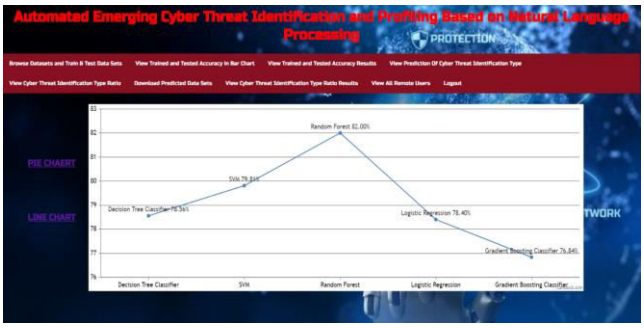

Fig 6.4 Accuracy in Line Chart

In above screen shows the accuracy in line chart.

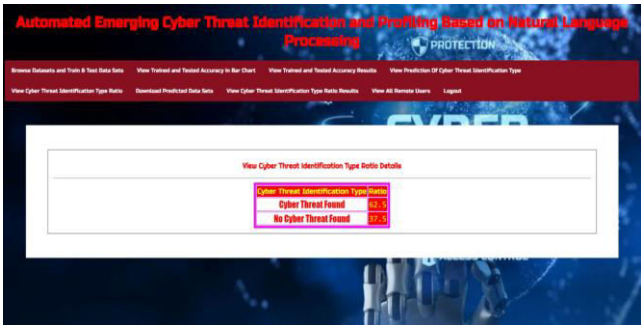

Fig 6.5 Cyber Threat Detection Ratio

In above screen shows the detection ratio of the cyber threats.

Fig 6.6 Cyber Threat Detection Ratio Graph

In above screen shows the detection ratio graph of the cyber threats.

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 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 (the subject).

8. 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. 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.
[21] B.-J. Koops, J.-H. Hoepman, and R. Leenes, ''Open-source intelligence and privacy by design,'' Comput. Law Secur. Rev., vol. 29, no. 6, pp. 676–688, Dec. 2013.
[22] R. Campiolo, L. A. F. Santos, D. M. Batista, and M. A. Gerosa, ''Evaluating the utilization of Twitter messages as a source of security alerts,'' in Proc. 28th Annu. ACM Symp. Appl. Comput., Mar. 2013, pp. 942–943.
[23] N. Dionísio, F. Alves, P. M. Ferreira, and A. Bessani, ''Cyberthreat detection from Twitter using deep neural networks,'' in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1–8.
[24] A. Niakanlahiji, J.Wei, and B. Chu, ''A natural language processing based trend analysis of advanced persistent threat techniques,'' in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2018, pp. 2995–3000.
[25] G. Ayoade, S. Chandra, L. Khan, K. Hamlen, and B. Thuraisingham, ''Automated threat report classification over multi-source data,'' in Proc. IEEE 4th Int. Conf. Collaboration Internet Comput. (CIC), Oct. 2018, pp. 236–245.