IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 8, Issue 1, 2020

SMARTER DETECTION OF NEW INTRODUCTORY CYBERTHREATS USING NATURAL LANGUAGE PROCESSING

Ms. S. Chandra Priyadharshini, Ms. R. Latha Priyadharshini, Mr. S. Satheesh, Mrs. A. Baranishri, Mr. S. Sugavanam

Associate Professor [1,5] Assistant Professor [2,3,4]

chandrapriyadharshini.s@actechnology.in, lathapriyadharshini.r@actechnology.in, ssatheesh@actechnology.in, baranishri.a@actechnology.in, sugavanam.s@actechnology.in

Department of CSE, Arjun College of Technology, Thamaraikulam, Coimbatore-Pollachi Highway, Coimbatore, Tamilnadu-642 120

ABSTRACT:

More and more, the amount of time that elapses between the discovery of a new cyber vulnerability and its exploitation by cybercriminals is shrinking. This is well shown by recent incidents, such the Log4j vulnerability. Attackers began searching the web for sites that could be susceptible to the exploit in order to install malware such as bitcoin miners and ransomware on those servers within hours of the vulnerability's announcement. For this reason, early threat and capability detection is crucial for cyber security defence strategies in order to maximise the efficacy of preventative measures. The enormous amount of data and information sources that need to be analysed for indications that a danger is growing makes finding new threats a tough undertaking for security analysts, despite how vital it is. To that end, we provide a system that can automatically detect and profile new threats based on their characteristics, with MITRE ATT&CK serving as a database of threat information and Twitter posts as an event source. The three key components of the framework are as follows: first, the ability to recognise and identify cyber threats; second, the use of two machine learning layers to filter and categorise tweets in order to create a profile of the detected danger; and third, the creation of alarms depending on the risk posed by the threat. In order to better understand the dangers and find ways to counter them, our study primarily focusses on a new way to profile them based on their intents or aims. Our tests showed that the profiling stage was 77% accurate in its threat profiling.

I. INTRODUCTION:

As the cyber landscape continues to evolve, the shrinking timeframe between the disclosure of vulnerabilities and their exploitation by threat actors presents a

ISSN 2347–3657
Volume 8, Issue 1, 2020

pressing challenge for cybersecurity. Recent incidents, exemplified by the Log4j vulnerability, vividly illustrate this trend. Within hours of its disclosure, malevolent entities swiftly initiated attacks, targeting vulnerable systems for deploying ransomware and cryptocurrency miners. This underscores the urgency for cybersecurity strategies to swiftly detect and comprehend emerging threats to maximize preemptive defense actions. Yet, amidst vast volumes of data, identifying nascent threats remains a formidable task for security analysts. To address this challenge, our project introduces a novel framework designed for the automatic identification and profiling of emergent cyber threats, utilizing Twitter as an event source and leveraging MITRE ATT&CK for threat characterization."

"The framework orchestrates three core components: first, the identification of cyber threats and their nomenclature; second, the profiling of these identified threats, discerning their intentions and goals through a sophisticated machine learning architecture; and finally, the generation of alerts based on the risk posed by the identified threats. A significant stride in our work lies in our approach to characterizing these emergent threats, providing contextual insights into their intentions. This added layer of understanding not only facilitates threat identification but also offers avenues for effective mitigation strategies. In our experimental endeavors, the profiling stage exhibited a commendable F1 score of 77%, demonstrating a robust capability in accurately profiling and understanding discovered threats."

"This project stands at the forefront of proactive cybersecurity measures, aiming to equip defenders with a sophisticated system capable of early threat detection and nuanced threat characterization. By leveraging Twitter as a valuable source of event data and employing cutting-edge machine learning techniques, the framework not only identifies threats but also delves deeper into their intentions, providing invaluable insights for proactive defense actions against rapidly evolving cyber threats.

III. 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

ISSN 2347–3657
Volume 8, Issue 1, 2020

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

ISSN 2347–3657
Volume 8, Issue 1, 2020

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.

In [26] an identification and classification approach of cyber threat indicators in the Twitter stream is presented. The research proposes a data-driven approach for
modeling and classification of tweets using a cascaded Convolutional Neural Network (CNN) architecture to both classify tweets as related or not to cyber security and classify the cyber-related tweets into a fixed listed of cyber threats. The proposed solution includes a pre-processing phase that uses IBM's Watson Natural Language API to identify tweets related to cyber security according to Watson classification results. Additionally, in the pre-processing phase, there is a pre-labeling step performed by simple string matching on the pure tweet text. The threat types considered were: ''vulnerability'', ''DDoS'', ''ransomware'', ''botnet'', ''data leak'', ''zero-day'' and ''general''. Further, the proposed approach uses CNN models trained to classify each tweet as relevant or irrelevant to cyber security. The relevant tweets are passed to a second CNN layer to be classified as one of the 8 different threat types mentioned above. There are important differences of our proposal compared to this one.

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

ISSN 2347–3657
Volume 8, Issue 1, 2020

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

ISSN 2347–3657
Volume 8, Issue 1, 2020

labels them as either relevant when the tweets contain security-related information, or irrelevant otherwise.

Relevant tweets are processed for information extraction by a Named Entity Recognition (NER) model, implemented as a Bidirectional Long Short-Term Memory (BiLSTM) neural network [30]. This network labels each word in a tweet with one of six entities used to locate relevant information. Furthermore, the authors chose to use the application of deep learning techniques because of its advantages in the NLP domain [31]. Thus, they propose an end-to-end threat intelligence tool that relies on neural networks with no feature engineering.

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.

IV. 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,

ISSN 2347–3657
Volume 8, Issue 1, 2020

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

V. 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

Login, Train & Test User Profile Data Sets, View User Profile Trained and Tested Accuracy in Bar Chart, View User Profile Trained and Tested Accuracy Results, View All Profile Identity Prediction, Find and View Profile Identity Prediction Ratio, View User Profile Identity Ratio Results, Download Predicted Data Sets, 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 profile identification status,view your profile.

VI. RESULT

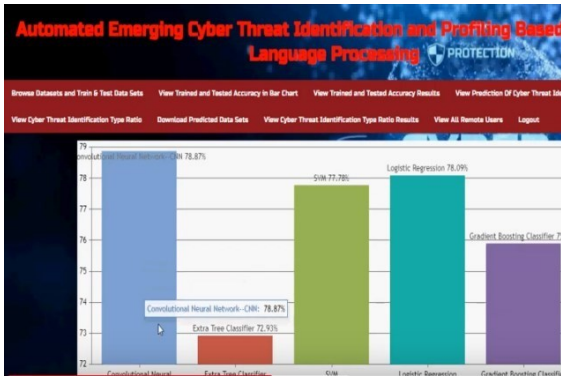
Integrate MITRE ATT&CK framework to enhance threat characterization and mapping. Leverage ATT&CK's structured knowledge base to correlate identified threats with known attack patterns, tactics, techniques, and procedures.

Continuously evaluate prediction models' performance using validation

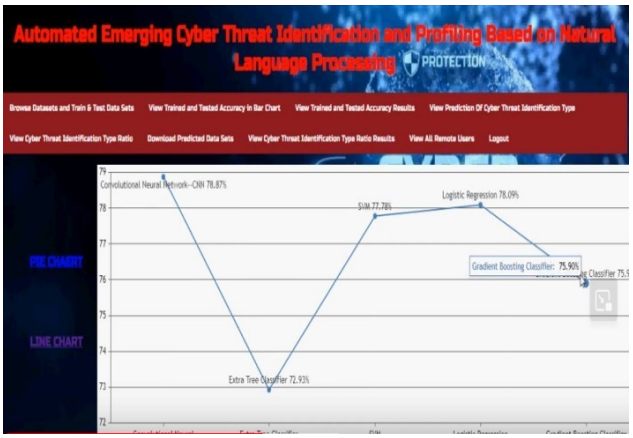
International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 8, Issue 1, 2020

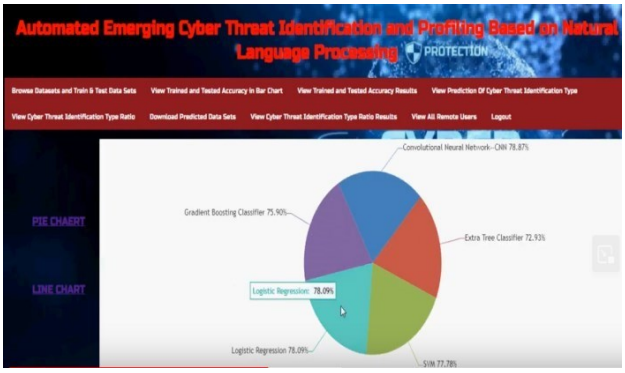
techniques and refine them based on feedback and new data to improve predictive accuracy.

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

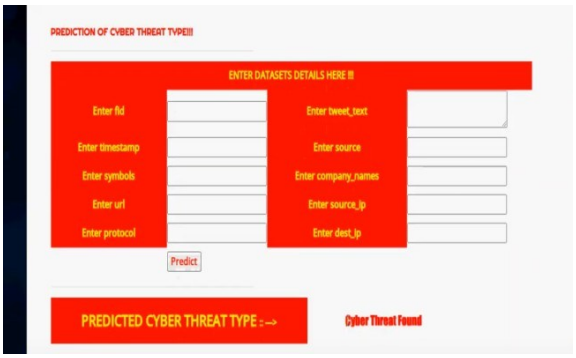


Line Charts Graph: Depict the trend of emerging threats over time, showcasing the frequency or intensity of threats detected.





Prediction of cyber threat type:



VI. CONCLUSION

ISSN 2347–3657
Volume 8, Issue 1, 2020

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

ISSN 2347–3657
Volume 8, Issue 1, 2020

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).

VII. 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*

ISSN 2347–3657
Volume 8, Issue 1, 2020

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.

ISSN 2347–3657
Volume 8, Issue 1, 2020

[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.