Cyber threat ontology and adversarial machine learning attacks: analysis and prediction perturbance

# Cyber Threat Ontology and Adversarial Machine Learning Attacks: Analysis and Prediction Perturbance

Abel Yeboah-Ofori[1] Umar Mukhtar Ismail[2], Tymoteusz Swidurski[3] Fransica Opoku-Boateng[4]
*School of Computer & Eng, Sch of Architecture, Computing & Eng, College of Computer & Cyber Science*
University of West London University of East London Dakota State University
London, UK London, UK Dakota, USA
abel.yeboah-ofori@uwl.ac.uk; u.ismail@uel.ac.uk; u1966166@uel.ac.uk, Francisca.Opoku-Boateng@dsu.edu

***Abstract:*** **Machine learning has been used in the cybersecurity domain to predict cyberattack trends. However, adversaries can inject malicious data into the dataset during training and testing to cause perturbance and predict false narratives. It has become challenging to analyse and predicate cyberattack correlations due to their fuzzy nature and lack of understanding of the threat landscape. Thus, it is imperative to use cyber threat ontology (CTO) concepts to extract relevant attack instances in CSC security for knowledge representation. This paper explores the challenges of CTO and adversarial machine learning (AML) attacks for threat prediction to improve cybersecurity. The novelty contributions are threefold. First, CTO concepts are considered for semantic mapping and definition of relationships for explicit knowledge of threat indicators. Secondly, AML techniques are deployed maliciously to manipulate algorithms during training and testing to predict false classifications models. Finally, we discuss the performance analysis of the classification models and how CTO provides automated means. The result shows that analysis of AML attacks and CTO concepts could be used for validating a mediated schema for specific vulnerabilities.**

***Keywords:*** *Adversarial Machine Learning; Cyber Threat Ontology; Threat Intelligence; Threat Prediction; Cybersecurity.*

## I. INTRODUCTION

Machine Learning (ML) techniques have been used on various algorithms to learn datasets [1][2] for classifications accuracies, analyzing network traffic, anomaly detection and threat predictions. However, adversaries are adopting AML strategies to inset malicious input data during training and testing dataset to cause perturbations and falsify the classification models. The application of ML techniques to cybersecurity is prone to inadequacies that limit its effectiveness to real-time cyberattack incident scenarios [4]. Adversarial machine learning techniques have been stealthily used to manipulate input data in order to exploit vulnerabilities in ML classification algorithms models to predict false classifiers [8]. Classification algorithms such as SVM and Neural Networks were considered robust to adversarial attacks [9, 10] however, recent AML attacks have proved that the classifiers are futile [11]. Some adversarial machine learning attacks involve adversaries using spam filtering techniques where misspelt words are inserted surreptitiously into the features to appear as legitimate words in spam messages [10] [12]. Thus, relying on ML predictions for threat intelligence gathering is inadequate and requires other methods of threat detections. Cyber threat an ontology considers the concepts of existing cyber threats becoming real and the phenomenon surrounding its existence. Several existing literature and methodologies consider ontology from information retrieval, knowledge representation, text mining and machine learning perspective that brings automation acquisition in ontology processing from unstructured data [13]. However, cyber threat ontology has not been given the shared conceptualizations and relationships required from the semantic web, cybersecurity, and adversaries machine learning perspective for knowledge acquisition. Ontology provides semantic mapping and defines the relationship between concepts to provide explicit knowledge and automated means for validating mediated schema for cyber threat intelligence (CTI) [13] [14]. Ontology from the cyber threat Intelligence perspective describes how attack concepts, properties, relationships, and their interdependencies are used in a formal and structured approach for threat analysis [15] [16]. The goal of the CTO is to extract relevant attack instances and information from data to ensure consistency and accuracy in security concepts and for knowledge reuse in the threat intelligence domain. CTO could facilitate knowledge reuse in threat intelligence for known attacks. ML works with a certain probability based on the analysed data and the decision the system will adopt. However, [5] demonstrating how deep neural networks could be manipulated with minor adjustments of the malicious input data is challenging.

This paper aims to explore Cyber Threat Ontology concepts and Adversarial Machine Learning Attacks for threat analysis and prediction to improve cybersecurity. The novelty contributions are threefold. Firstly, cyber threat ontology concepts are considered to model advance persistent threat (APT) attack. Secondly, we discuss how adversarial machine learning techniques are used on a dataset through malicious input to present a false narrative. Finally, a performance analysis is carried out on the classification algorithms for the predictions and how CTO and AML can improve security. The result shows that analysis of AML attacks and CTO concepts could be used for validating a mediated schema for specific vulnerabilities.

## II. RELATED WORKS

This section discusses the related works in cyber threat ontology and adversarial machine learning relevant to the cybersecurity domain.

### *A. Cyber Threat Ontology*

Cyber threat ontology from a security perspective describes organizational security concepts, properties relationships, and interdependencies in a formal and structured approach for analysis and intelligence gatherings [13] [14]. The goal of cyber threat ontology considers the extraction of relevant attack instances and threat information from data to ensure consistency and accuracy in the cybersecurity concepts for knowledge reuse in the threat intelligence domain. For instance, Asim et al 2018 did a survey on various ontology learning techniques and applications by highlighting and evaluating the pros and cons and discuss the different algorithms [13]. Gao et al. 2013, proposed an ontology-based model of network and computer attacks for security assessment and standards classifications that establishes relationships among network security services, threats, vulnerabilities and causes of failures [14]. Gyrard et al. 2013, proposed an ontology for security toolbox, attacks and countermeasures from a secure e-governance applications perspective for capturing and presenting concepts of security requirements in application development of security expert knowledge [15]. Herzog et al. 2007, proposed an ontology-driven approach of information security concepts for analyzing and sharing intelligence vocabularies [22]. Hu et al. 2012, proposed an ontological approach to information security education from an OWL-based security incident that defines security incidents using unique vocabularies for the concepts and their associated relationships of various incidents and information management and sharing [17]. Mozzaquatro et al 2018 proposed an ontology-based cybersecurity framework for the Internet of Things that considers design time and provides a dynamic method to build security and run time that monitors the IoT environment for analysis [18]. [19]. Jia et al 2017 proposed a practical approach to constructing a knowledge graph for cybersecurity by using machine learning to extract entities and building ontologies to obtain cybersecurity based knowledge [20] and security ontology with model driven architecture for software development [21].

### *B. Adversarial Machine Learning Attacks*

The adversarial machine learning technique is used by the adversary to inject malicious input data in the dataset during the training and testing phase to manipulate the classification model [12]. The method is used in supervised learning algorithms for cybersecurity datasets to exploit vulnerabilities and compromise performance results [8] such as spam filters and IDS/IPS when predicting cyberattack trends and predicting probability of fraudulent activities. The adversary could cause an increase in the false-positive rates by inserting malicious samples in the test phase to generate wrong classifications rates of the sample data. The adversarial machine learning technique could be used to manipulate training data to violate security policy, gain knowledge of threat intelligence, adversary capabilities and level of manipulations [22]. Apruzzese et al 2019 applied adversarial attacks on random forest, Multi-layer perception, and K-Nearest Neighbour classifiers. [12]. Kravchik et al 2020 proposed evasive and poisoning attacks on cyberattacks detectors for industrial control systems by using Neural network-based methods and backwards-gradient based poisoning [23]. Duddu 2018 examined AML attacks by considering various techniques for adversarial modelling. And used a testbed for the analysis [11]. Zhang et al 2015 proposed a novel adversary-aware feature selection model using wrapper-based feature selection on linear SVM, perception and non-linear classifiers with forward selection and backward elimination [24]. Biggio et al 2011 explore adversarial data manipulations using SVM classification algorithm under adversarial label noise by subverting the SVM learning process [20]. Chen et al 2017 designed a randomizing SVM model by using robust SVMs Gaussian distribution against method for adversarial attacks [26]. Munoz-Gonzalez et al 2017 proposed a novel poisoning algorithm based on a back-gradient optimization and used neural networks algorithms and deep learning architectures techniques to learn the dataset. [22]. Jagielski et al. 2018 introduced a systematic study of poisoning attacks on linear regression model by using standard-gradient algorithm and baseline gradient [24].

The related works are all relevant and contribute towards the improvement of the knowledge of AML attacks. However, none of the works considers the applying cyber threat ontology concepts and adversarial machine learn approach for improving cybersecurity threat analysis and predictions.

## III. APPROACH

This section provides an overview of the proposed approach from conceptual and ontological perspectives and the process used for the analysis and prediction of AML attacks.

### *A. Rationale for Implementing CTO and AML*

The rationale for implementing cyber threat ontology and adversarial machine learning is based on the premise that the cyberattack phenomenon includes many uncertainties that make the threat landscape unpredictable. Additionally, due to the varying organizational goals and dynamic system requirements, various integration, varying business processes and delivery mechanisms, predicting cyberattacks from an organizational setting perspective has been challenging. To address these problems, we consider CTO concepts for knowledge representation and the AML attack approach to predict false narrative. Therefore, the main rationale for using this method is:

- First, we model CTO for advance persistent threat (APT) attacks for threat mapping and the properties to determine the causal relationships for knowledge representation.

- Secondly, we apply AML attack on the dataset through malicious insertion to predict false narratives cyberattacks. We follow Figure 1 for the CTO and AML approach.

### B. *Cyber Threat Ontology for Knowledge Representation*

Cyber threat ontology is considered as key to successful knowledge representation, semantic visualization and reuse of critical knowledge, especially for threat analysis [17]. We consider Advanced Persistent Threat attacks modelling using attack phases such as infiltration, manipulation, exfiltration, and obfuscation. The ontologies provide other benefits that can be used to consolidate and clarify the definition, attributes, and relationship between concepts to eliminate the vagueness and ambiguity of threat knowledge among actors while also facilitating consistent elicitation of relevant controls. Additionally, by the explicit representation of knowledge, ontologies can be used to form solid knowledge threat pattern, prevent, detect, and respond to threats. For instance, by providing a coherent and formal representation of threat actors, an ontology provides a common language. The tactics, technique, and threat procedure can be easily shared amongst all actors in the domain. Ontology also presents concepts, properties, relationships, and interdependencies in a formal and structured approach [16]. The process includes extracting relevant attack instances and threat intelligence from data to ensure consistency and accuracy in the security domain. To address trust and information assurance issues, organizations need to map their security relationships, dependencies, and vulnerabilities inclusively. Ontology also uses insider threat indicators to provide a common language with which to represent and share knowledge and consistently model indicators of insider threats.

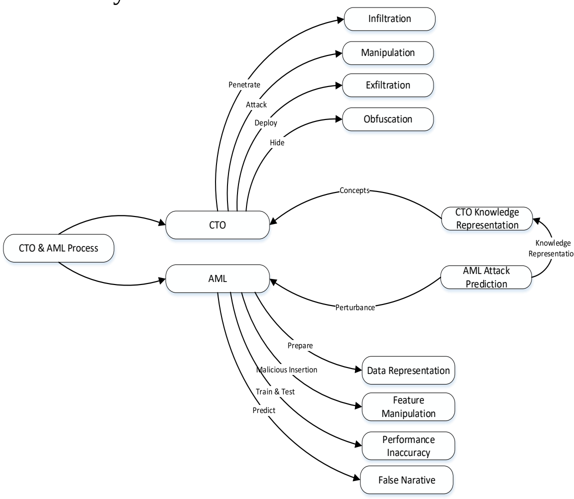

Fig. 1. Cyber Threat Ontology and Adversarial Machine Learning

### C. *Adversarial Machine Learning for Prediction*

Adversarial Machine learning (AML) are malicious attacks deployed to manipulate features to cause perturbation to the data during training and testing to predict false narratives. ML classification algorithms use supervised and unsupervised techniques to learning the dataset depending on the performance requirements [12]. Supervised learning supports classification and regression test for performances accuracy during training and testing. Unsupervised learning performs well on ancillary tasks such as data clustering [22]. We used a supervised learning approach to learn the dataset. In supervised learning, the AML techniques use malicious input to attempt to deceive the classification models during training and testing.

## IV. IMPLEMENTATION

This section follows our approach for the cyber threat ontology (CTO) and AML techniques to our implementation. To achieve the applicability of our work, we use the CTO learning to describe the security concepts, properties and the relationships required to model a security goal. The AML techniques are applied on a dataset through malicious insertion to cause perturbance in the features for the misclassifications.

### A. *Cyber Threat Ontology Using Advanced Persistent Threat*

The cyber threat ontology implementation considers a conceptual model to identify and map the concepts that drive the required entities, properties relationships and rule sets for the cyberattack domain. The concepts include infiltration, manipulation, exfiltration, and obfuscation as well as the properties that provide the conceptual reasoning, relational knowledge and understanding of cyber threat intelligence required. We implement the CTO for advanced persistent threat attack phases using the four key steps as shown in Figure 2.

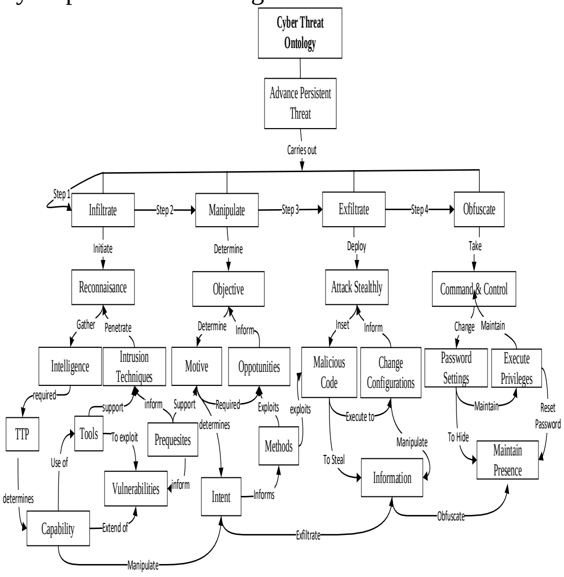

Fig. 2. Cyber Threat Ontology Conceptual Model Using Advanced Persistent Threat Attack

### A. *Advance Persistent Threat (APT) Attack*

APT is a type of cyberattack that the attacker uses various attack methods including reconnaissance to penetrate a network system, gain access and stealthy steal information and take command and control of their victims' system. We follow the four key steps and the implementation process including infiltration, manipulation, exfiltration, and obfuscation.

- **Step 1: Infiltration**:- The infiltration step determines the information the attacker requires to penetrate the system. That includes carrying out reconnaissance for intelligence gatherings of the victim's network to identify the overall organizational assets, including hardware, software, network infrastructure, design process and policy uses in the environment. That includes internal, external, and third-party vendors on the network. The intelligence gathered informs the adversary of the tactics, techniques, and procedures (TTP) to de deployed, the capabilities required, the toolsets, the vulnerable spots, the prerequisites for the attack and the intrusion techniques to penetrate the network. The attacker could use remote access trojan (RAT) to gain access for the penetration.
- **Step 2. Manipulation:-** The manipulation step follows the infiltrations phase after penetrating the systems. It considers the objective of the cybercrimes and determines the attacker's motive and intents of the attack. The attack method including the TTP is used to exploit the vulnerable spots on the network, software, and system users depending on the opportunities available to the attacker. The intent of the attacker could be to commit cybercrimes such as Intellectual Property theft, ID theft, Data theft and Industrial Espionage attacks.
- **Step 3: Exfiltration: -** The exfiltration step considers how the attacker deploys stealthy attacks to syphon data after penetration and manipulating the system network and software. The approach could be to insert a malicious code that provides a backdoor to steal information and change configurations for continuous manipulations.
- **Step 4: Obfuscation: -** The obfuscation step aims at taken command & control and maintaining a presence as the final phase of the attack after the infiltration, manipulation, exfiltration phases. The APT attacker usually takes command & control of the system, and based on the capability of the attacker, the attack can change the password settings as normal users, escalate movement, maintain lateral movement, execute privileges and reset password to maintain a presence.

The steps provide us with the ontological view from an APT attack perspective and the knowledge representation of the CTO required for mapping the formal language. The ontological view enables the explicit specification and conceptualization of ideas representing an abstract model of the phenomenon. The use of CTO for APT attack concepts has enabled the construction of knowledge representation in organized metadata of complex information regarding security.

## B. *Determining the Performance Accuracies During Normal Training and Testing Time*

Figure 3 looks at the results of the accuracies of combining 2 classification algorithms RF and GB in a pipeline and run in a ROC curve to determine the true positive and false positive rates using the 10-Fold cross-validation. RF produces a performance result of 73% compared to GB 79% with a majority voting of 78%. The highest classifier from the performance model was GB as it can predict better performance in predicting attack. However, the results show a slight reduction in the overall score with the MV score of 78%. Further, it shows higher accuracy for the TPRs and FPRs as compared to figure 2 where the performance went down when we included the RF algorithm.

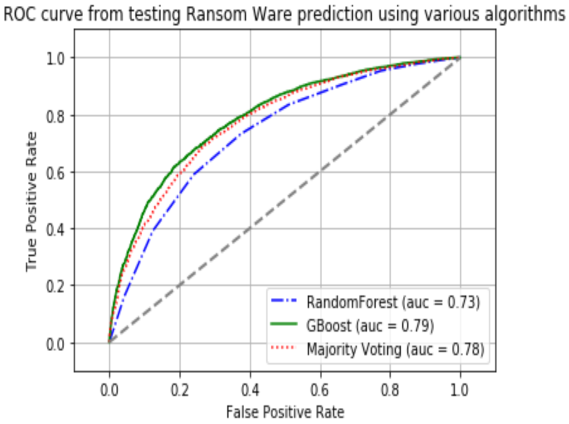

Fig 3. Roc Curve for Prediction the RF and GBoost Algorithms in MV

## C. *Adversarial ML Attack Modelling*

The AML attack experimentation considers different classification models for detecting the integrity and violated (non-integrity) attacks. The integrity attacks are those performed during training and testing time on a normal dataset with correct features. The violated attacks are those performed on malicious data to caused perturbance through feature manipulations during the learning phase. The experiment is based on two supervised classification algorithms RF and GBoost. We consider malicious attack modelling, then use a testbed for the experiment and the performances accuracies. The performances are evaluated in adversarial machine learning scenarios for predictive analytics. The attack modelling distinguishes the integrity attack model and the violated malicious attack model. The classification of the AML attacks considers the attacker goal, knowledge, capability, and strategy to assist in understanding the motive and intent of the attacker. A goal represents what an adversary may want to achieve which is to violate the integrity, availability, and confidentiality of the security system. The Knowledge represents the intelligence gathered by the attacker using reconnaissance and the ML algorithms, features, and datasets. The capability considers actions that an attacker could maliciously deploy to exploit various instances of the features during the learning phases to cause perturbance and false narratives. The strategy considers the procedures adopted by the adversary to pursue the intents by leveraging on the information gathered through the reconnaissance and exploits [22]. The testbed for the adversarial attack is implemented on a collection of publicly available datasets by Microsoft Windows Defender [28]. The dataset has over 40,000 entries with 62 columns and each row represents different telemetry data entries. Each row in the dataset corresponds to a machine uniquely identified by a machine Identifier. The overall features were 64. We extracted 38 features in the primary data relevant to the attack profile [26] [27].

### D. *Adversarial Machine Learning Algorithm*

We provide an algorithm for the adversarial machine learning attack that considers RF and GBoost classifiers in a multiclass. The data (d) was collected from [28]. We consider attack (X) as the input and a class for each hyperplane. The classifier (C) is feed into the feature (f). The malicious attack (X) class is to locate the algorithms and push them beyond the nearest hyperplane to cause the perturbance (p) and possible misclassification in the performance. We use the following algorithm to explain the steps [30].

**Algorithm for Adversarial Attack on RF & GBoost Classifiers**

```
1.  Input: Cyberattack $X$, Classifier $C$.
2.  Output: Perturbance $p$.
3.  Identify the feature: $f$
4.  Initialize $X_0 \leftarrow X, i \leftarrow 0$
5.  While $p(X_0) = p(X_0)$ do
6.      For $p \neq p(X_0)$ do
7.          $W^{|}_k \leftarrow f_k (X_i) - f_k (X_0)^{(xi)}$
8.          $C^{|}_k \leftarrow f_k (X_i) - f_{k(x0\}}(X_0)$
9.  end for
10. $I \leftarrow \arg\min_{p \neq p(x0)} \underline{|c^{|}_k| /} \| c^{|}_k\|_2$
11. $r_i \leftarrow |c^{|}_k| / \| c^{|}_k\|^2_2 \;\; W^{|}_k$
12. $X_i+1 \leftarrow X_i + r_i$
13. $i \leftarrow I + 1$
14. end while
15. return $r = \sum_i r_i$
```

**Adversarial Attack Algorithm Pseudo Code**

1. Start input attack (X) and the classier (C) that determines the model
2. The Output is the perturbance (p) that indicates the malicious insertion
3. Set the features (f) space that will be corrupted
4. Initialize the perturbed dataset (d) in the loop variable with the original attack
5. Start the iteration and continue until the original attack and perturbed attack are not equal
6. Consider the number of attacks (X) classes with the most probability after the original attack.
7. Store minimum difference between original gradient with each attack (XkXK) classifiers.
8. Store the difference in the labels (CkCK)
9. End Loop
10. Store inner loop with minimum XkXK and CkCK. Use that to calculate the closest hyperplane or input (X)
11. Calculate minimal vector that project (X) onto closest hyperplane that was calculated in 10
12. Add minimal perturbance to the dataset (d) and check if it misclassified.
13. Increase the loop variable.
14. End loop
15. Return the total perturbance (p) that indicates the overall sum of the calculated perturbances.

### D. *Adversarial Machine Learning Attacks During Training Time*

The adversarial machine learning attack experiment considers predicting a ransomware attack that involves malicious inserting an input into the features to violate the integrity of the performance during training and test time. We consider an attack model where the adversary has penetrated the network using remote access trojan attack and can manipulate the system using advance persistent threats (ATP) and command and control (C&C) capabilities. The adversary goal is fourfold: Infiltrate, Manipulate, Exfiltrate, and Obfuscate. First, the adversary infiltrates the network server by penetrating through the network nodes using remote access trojan (RAT) attack to gain access. Secondly, the attacker uses command and control capabilities, manipulates the system, and insert a malicious code to modify the network intrusion detection tools that could propagate to other networks. Thirdly, the attacker exfiltrates information, causing cybercrimes such as Industrial Espionage attack, ID theft, Intellectual Property theft and Ransomware attack. Finally, the attacker obfuscates through the system, changes his password regularly, and hide in the systems. The objective of the adversarial attack is to modify the network intrusion detection by inserting input data into the dataset to increment the features during training and testing time to predict false narratives. We consider the following scenarios for the implementation:

- Scenario 1: Predicting the accurate responses of the FR and GBoost classifiers based on cyberattack type.
- Scenario 2: Predicting Accuracies after Adversarial Machine Learning deployed based responses cyberattacks.

Scenario 1, we use ML techniques on FR and GBoost classification algorithm to learn dataset for performance accuracies and analyse the predictions for accurate responses based on the type of cyberattacks.

**In scenario 2**, we simulate the adversarial attack scenario by altering the features by inserting the script in the features that will exchange the duration of the bytes in the total packets during learning. The objective is to cause perturbance during training and testing time to predict the false narratives.

## V. RESULTS

This section analyses the results on the adversarial attacks scenarios using ML algorithms and presents the performance accuracies of the different threats that were proposed during training and testing for the FR and GBoost classification algorithms. The results are determined using the Precision, Recall and F-Score.

- **Scenario 1: Predicting attack responses of the FR and GBoost classifiers based cyberattacks.**

The performance predicts the accurate responses of the FR and GBoost classification algorithms based on the type of cyberattacks. Table 1 presents the performance of the classifications of the RF and GBoost algorithms in identifying the multiple responses of cyberattacks based on the given malicious attack. From Table 1, RF achieved an accuracy of 73% and GBoost 78%. Comparing the performance of the classifiers, GBoost performed better for precision, recall and F-score, whilst RF received a low Precision (P), Recall (R) and F-Score (F).he attack's categories indicates that Malware, Ransomware and spyware attacks provided different responses with 79% accuracy

Table 1. Predicting cyberattacks responses on Endpoint Nodes Using RF and GBoost Classifiers

| ACCURACY | RF 72% | | | GBOOST 79% | | |
|---|---|---|---|---|---|---|
| ATTACKS | P | R | F | P | R | F |
| Ransomware | 0.73 | 0.71 | 0.72 | 0.78 | 0.76 | 0.77 |
| RAT | 0.69 | 0.66 | 0.67 | 0.72 | 0.70 | 0.71 |
| Malware | 0.72 | 0.69 | 0.72 | 0.79 | 0.76 | 0.77 |
| Spyware | 0.73 | 0.70 | 0.71 | 0.78 | 0.75 | 0.76 |

- **Scenario 2: Predicting the responses after inserting adversarial attacks.**

Scenario 2 considers the performances classifiers after the adversarial attack sample is deployed on the classifies. A malicious script is inserted to alter the features during the retraining of the dataset. We inset the script to exchange the bytes duration for the total packets into the features during learning. We tested the RF and GBoost classifiers again on the dataset. The results in Table 2 show that the inputted script changed the features of the training and testing, reduced the data and invalidate the effects of the performance. The adversarial attack changed the dataset as that does not include features related to the duration, exchange bytes and the total packets. The goal is to affect the duration of the runtime during the training and testing which will then impact the exchanged bytes since the duration is minimal and consequently reduce the total packets that are being retrained and retested due to the duration. The response of the performance accuracies after adversarial attacks has been deployed is determined based on the responses of the various accuracies on the cyberattacks during training and testing time. Table 2 presents the effects of the performance of the classifiers after the insertion of the adversarial attack based on the cyberattacks. The malicious insertion severely affected the percentage rates of the classification algorithms during the retaining and reduces the figures to about a third due to the effect of the changes in the feature selections.

Table 2. Compares the differences after malicious adversarial attack

| ACCURACY | RF 22% | | | Gboost 25% | | |
|---|---|---|---|---|---|---|
| ATTACKS | P | R | F | P | R | F |
| Ransomware | 0.24 | 0.20 | 0.22 | 0.26 | 0.24 | 0.25 |
| RAT | 0.22 | 0.69 | 0.21 | 0.25 | 0.23 | 0.24 |
| Malware | 0.24 | 0.21 | 0.23 | 0.27 | 0.24 | 0.25 |
| Spyware | 0.23 | 0.21 | 0.22 | 0.26 | 0.24 | 0.25 |

Table 2 compares the differences in the performances between the classifiers before and after the insertion of the malicious adversarial attack. The decree in the prediction after the malicious attack shows the effect of the perturbations after the adversarial attack on the classifiers during retaining. It should be noted that the feature removal technique adopted to cause the perturbance may not be applicable to other adversarial attacks. Comparing the performances in Table 1 and 2, the RF achieved an accuracy of 22% and  GBoost achieved a higher accuracy of 25% for the precision, recall and F-score respectively. Furthermore, the ransomware, malware, and spyware attacks identified different responses with an average of 25% accuracy for the harmonic mean RAT attack.

## VI. DISCUSSION

Predicting cyberattack trends has been challenging as defending against future cyber threats relies on machine learning techniques to learn dataset for performance accuracies and detection. Adversaries are exploiting these techniques by maliciously poising datasets to subvert classification performances to predict false narratives. The effects of the poison attacks on a dataset focused on the integrity violations on the performance accuracies. The adversary goal was fourfold: Infiltrate, Manipulate, Exfiltrate, and Obfuscate. The adversarial attack model considered a penetrated network using remote access trojan (RAT) attack. The adversarial attacker was able to misclassify the prediction using advance persistent threats (ATP) and command and control (C&C) capabilities. The paper measured the effectiveness of malicious attacks on the feature selection criteria on the dataset during the retraining and retesting time to determine the degree of the poisoned rates on the ransomware, malware, spyware, and RAT attacks.  Further, we compared the performance accuracies of the classifiers during the normal learning of the dataset to the retrained one. The results in Table 2 indicated a significant drop in the rates of performances on the scores compares to Table 1 depending on the type of cyberattacks after the malicious insertion.  The goal of analysing the CTO for the APT attack was to extract relevant attack instances from the intelligence for accuracy in security concepts and knowledge reuse in the threat intelligence domain. The CTO acted as a link that connects the concepts with the threat information required for predicting future trends.

## VII. CONCLUSION

The paper had revealed several challenges in ML predictions as adversaries are executing arbitrary commands maliciously to manipulate data. The paper has discussed the relevance of developing cyber threat ontology concepts from the advanced persistent threat perspective. That describes how attack concepts, properties, relationships, and their interdependencies are used in a formal and structured approach for threat analysis and knowledge reuse. The cyber threat ontology model extracted relevant attack instances and information for accurate and consistent security concepts and knowledge reuse in the threat intelligence domain. The cyber threat ontology model of APT attack provided a security assessment and attack classifications that establish relationships among threats, vulnerabilities and attack instances. Due to the invisibility nature of cyberattacks, the application of cyber threat ontology enables the exchange, sharing and reuse of cyber threat information on machine learning threat predictive analytics for security control mechanisms.  Future works include will consider AML attacks on other classifiers and datasets from other sources.

## REFERENCES

[1] A. Yeboah-Ofori and C. Boachie, "Malware Attack Predictive Analytics in a Cyber Supply Chain Context Using Machine Learning," *2019 International Conference on Cyber Security*

*and Internet of Things (ICSIoT)*, 2019, pp. 66-73, doi: 10.1109/ICSIoT47925.2019.00019.

[2] A. Yeboah-Ofori. "Classification of Malware Attacks Using Machine Learning in Decision Tree." International Journal of Security. (IJS). Vol. 11 (Issue-2), Pages 10-25. 2020.

[3] B. Biggio, F, Roli, "Wild patterns: Ten years after the rise of adversarial machine learning". Pattern Recognition. 84: 317–331. 2018. arXiv:1712.03141. doi:10.1016/j.patcog.2018.07.023.

[4] B. Biggio, G. Fumera, F. Roli,. "Multiple classifier systems for robust classifier design in adversarial environments". International Journal of Machine Learning and Cybernetics. **1** (1–4): 27–41. 2010. doi:10.1007/s13042-010-0007-7. ISSN 1868-8071.

[5] A. Kurakin, Samy; I. J. Goodfellow, S. Bengio,; (2017). "Adversarial Machine Learning at Scale". 2018.AI. arXiv:1611.01236.2016arXiv161101236K.

[6] V. Duddu. "A Survey of Adversarial Machine Learning in Cyber Warfare." Defence Science Journal, 68(4), 2018356-366. https://doi.org/10.14429/dsj.68.12371

[7] G. Appruzzese, L. Ferretti, M. Marchetti, M. Colajanni, A. and Guido, "On the Effectiveness of Machine Learning for Cyber Security. International Conference on Cyber Conflict." 2018. IEEE. doi: 10.23919/CYCON.2018.8405026

[8] M. N. Asim, M. Wasim, M . U. G. Khan, W. Mahmood, H. M, Abbasi, "A survey of ontology learning techniques" and applications, *Database*, Vol 2018 1-24, 2018, bay101, https://doi.org/10.1093/database/bay101

[9] J. Gao, B. Zhang, X. Chen and Z. Luo. "Ontology-based model of network and computer attacks for security assessment". Journal. Shanghai Jiaotong Univ. 18(5):554–562, 2013. DOI: 10.1007/s12204-013-1439-5

[10] A. Gyrard, C. Bonnet and K. Boudaoud. "The STAC (Security Toolbox: Attacks & Countermeasures) Ontology". In Proceedings of the 22nd international conference on World Wide Web companion, pages 165–166, Brazil, 2013.

[11] A. Herzog. N. Shahmehri and C. Duma. "An ontology of information security", International Journal Information Security. Priv. 1(4):1–23, 2007.

[12] H. Hu, M. Yang, Y. Ge, H. Xiang, and L. Fu, "An Ontological Approach to Information Security Education". In Proceedings of the 2nd International Conference on Future Computers in Education pages 160–165. China, 2012.

[13] B. A. Mozzaquatro, C. Agostinho, D. Goncalves, J. Martins, R. Jardim-Goncalves. "An Ontology-Based Cybersecurity Framework for the Internet of Things." MDPI. *Sensors*. 18(9):3053. 2018; https://doi.org/10.3390/s18093053.

[14] A. Ekelhart S. Fenz , A. M. Tjoa, E. R. Weippl . "Security Issues for the Use of Semantic Web in E-Commerce. In: Abramowicz W. (eds) Business Information Systems. Lecture Notes in Computer Science, vol 4439. Springer, Berlin, Heidelberg. 2007. https://doi.org/10.1007/978-3-540-72035-5_1.

[15] Y. Jia, Y. Qi, H. Shang, R. Jiang, A. Li, "A Practical Approach to Constructing a Knowledge Graph for Cybersecurity, Engineering." Volume 4, Issue 1, pp 53-60, 2018, https://doi.org/10.1016/j.eng.2018.01.004.

[16] W. Kang and Y. Liang. "A Security Ontology with MDA for Software Development". In Proceedings of the 2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. pp 67–74, 2013.

[17] B. Biggio, B. Nelson, P. Laskov. "Poisoning Attacks Against Support Vector Machine." ICML'12: Proceedings of the 29th International Conference on International Conference on Machine Learning. ACL. Pages 1467–1474. 2012.

[18] M. Kravchik, B. Biggio, A. Shabtai. "Poisoning Attacks on Cyber Attack Detectors for Industrial Control Systems" Cornell University. 2013. arXiv:1206.6389v3.

[19] F. Zhang, P. P. K. Chan, B. Biggio, D. S. Yeung and F. Roli, "Adversarial Feature Selection Against Evasion Attacks," in *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 766-777, March 2016, doi: 10.1109/TCYB.2015.2415032.

[20] B. Biggio, B. Nelson, P. Laskov. "Support Vector Machines Under Adversarial Label Noise" JMLR: Workshop and Conference Proceedings. Asian Conference on Machine Learning 20. 97–112. 2011.

[21] Y. Chen, W. Wang, X. Zhang X. "Randomizing SVM Against Adversarial Attacks Under Uncertainty." Advances in Knowledge Discovery and Data Mining. PAKDD. vol 10939. Springer, 2018. https://doi.org/10.1007/978-3-319-93040-4_44.

[22] L. Muñoz-González, B. Biggio, A. Demontis, A. Paudice, V. Wongrassamee, E. C. Lupu, F. Roli. "Towards Poisoning of Deep Learning Algorithms with Back-gradient Optimization." Cornell University. 2017. arXiv:1708.08689v1.

[23] Microsoft Malware Prediction, Research Prediction. 2019. Available: https://www.kaggle.com/c/microsoft-malware-prediction/data.

[24] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru and B. Li, "Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning," *2018 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA, 2018, pp. 19-35, doi: 10.1109/SP.2018.00057.

[25] A. Chakraborty "Introduction to Adversarial Machine Learning". FloydHub. 2029. Online. Available: https://blog.floydhub.com/introduction-to-adversarial-machine-learning/.

[26] A. Yeboah-Ofori, U. Ismai, S. Islam, H. Mouratidis, and S. Papastergiou. "Cyber Supply Chain Threat Analysis and Prediction using Machine Learning and Ontology". In: Maglogiannis I., Macintyre J., Iliadis L. (eds) Artificial Intelligence Applications and Innovations. AIAI 2021. IFIP Advances in Information and Communication Technology, vol 627. Springer, Cham. https://doi.org/10.1007/978-3-030-79150-6_41

[27] A. Yeboah-Ofori *et al*., "Cyber Threat Predictive Analytics for Improving Cyber Supply Chain Security," in *IEEE Access*, doi: 10.1109/ACCESS.2021.3087109.