Securing Intelligent Autonomous Systems Through Artificial Intelligence

Ganapathy Mani[a], Bharat Bhargava[a], Jason Kobes[b], Justin King[b], James MacDonald[b]

[a] *Purdue University, West Lafayette, Indiana, USA*
[b] *Northrop Grumman Corporation, McLean, Virginia, USA*

Abstract

Intelligent Autonomous Systems (IAS) reconstruct their perception through adaptive learning and meet mission objectives. IAS are highly cognitive, rich in knowledge discovery, reflective through rapid adaptation, and provide security assurance. It is paramount to have effective reasoning, decision-making, and understanding of operational context since IAS are exposed to advanced multi-stage attacks during training and inference time. Advanced malware types such as file-less malware with benign initial execution phase can mislead IAS to accept them as normal processes and execute malicious code later. IAS are also exposed to adaptive poisoning attacks where adversary inputs malicious data into training/testing set to manipulate the learning. Hence it is vital to monitor IAS activities/interactions to conduct forensics. This project will advance science of security in IAS through multifaceted advanced analytics, cognitive and adversarial machine learning, and cyber attribution based on the following approaches.

(a) Implement deep learning-based application profiling to categorize adaptive cyber-attacks and poison attacks on machine learning models using contextual information about the origin, trust, and transformation of data.
(b) Using HW/OS/SW data to develop perception algorithms using LSTM deep neural networks for detecting malware/anomalies and classifying dynamic attack contexts.
(c) Facilitate cyber attribution for forensics through privacy-preserving provenance structure for knowledge representation and perform intrusion detection sampling on HW /OS/SW data.
(d) Employ advanced data analytics to aid ontological and semantic reasoning models to enhance decision-making, attack adaptiveness, and self-healing.

Keywords [1]

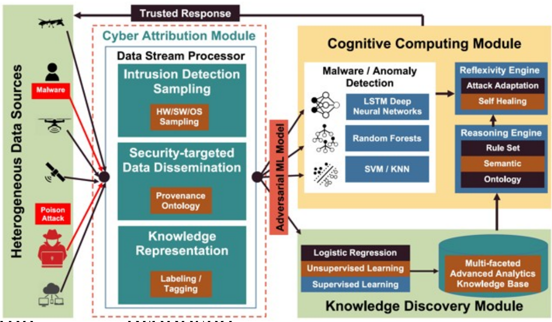
autonomy, machine learning, deep learning, cybersecurity, lstm

1. Solution Overview

Our focus is on constraints, barriers and challenges such as poorly understood attack surfaces, data set training availability and biases, processing latency, human understanding of AI results, AI/ML countermeasures, human-machine disparity, measurement of effects. We propose novel approaches for privacy-preserving cyber attribution, intrusion detection, adversarial

[1]Proceedings Name, Month XX–XX, YYYY, City, Country
EMAIL: email1@mail.com (A. 1); email2@mail.com (A. 2); email3@mail.com (A. 3)
ORCID: XXXX-XXXX-XXXX-XXXX (A. 1); XXXX-XXXX-XXXX-XXXX (A. 2); XXXX-XXXX-XXXX-XXXX (A. 3)

machine learning, malware/anomaly detection, reasoning, and decision-making. Cyber attribution involves extracting software, hardware, and operating system data to perform intrusion detection sampling (fixed or dynamic sampling), generating efficient provenance structure that is populated with specific data required for a particular analysis or learning, and labeling and tagging to properly represent the information obtained. The processed data is distributed to the cognitive module where the data is checked for any malicious data presence through poison attack filter. The filtered data is transmitted to cognitive computing module and knowledge discovery module, where the data is fed into supervised, unsupervised, and LSTM models to perform learning and advanced analytics. Based on multifaceted dimensions of data analytics, reasoning and decision-making ability of IAS are enhanced. The overall architecture of the proposed model-secure intelligent autonomous systems with cyber attribution-is demonstrated in figure 1.



the proposed unified architecture are given as follows:

- Intelligent autonomous systems receive large amounts of diverse data from various data sources. In addition, they operate in a dynamic operational context and interact with numerous entities such as other TAS, UAVs, satellites, sensors, cloud systems, analysts, malicious actors, and compromised systems.
- Cyber attribution module constitutes a stream data processor where data streams are

labeled / tagged on-the-fly for better knowledge representation and categorization. This data is stored as monitored or provenance data with its origin and historical information. For preserving privacy, detailed provenance data is reduced in its scope to include only necessary data for a particular analysis or learning. This module uses Provenance Ontology (PROV-O) structure (elaborated in a later section) to obscure unnecessary or privacy-compromising data. Furthermore, the attribution model monitors data generated by software (application parameters), hardware (memory bytes and instructions), and operating system (system calls). This data is used to conduct periodic sampling to identify signatures of intrusion activities.

- Once the data is processed, it goes through adversarial machine learning model. Attackers can insert malicious data into training and testing dataset to influence machine learning models. In order to isolate poisonous data, poison data filter performs methods such as classification of verified and unverified data as well as outlier extraction. Once the poisonous data is removed the data (raw or provenance data) is sent to Cognitive computing module and Knowledge discovery module.
- In Cognitive computing module, depends on the data and efficiency of machine learning methods, malware / anomaly detection is

performed through either deep learning methodologies such as Long short-term memory (LSTM) e.g. Recurrent Neural Networks (RNN) or Convolutional Neural Networks (CNN) or light-weight yet powerful machine learning methods such as Support Vector Machines (SVM), Random Forests (RF), and K-Nearest Neighbors (KNN). In addition, cognitive computing module consists of reasoning engine, which is driven by rule sets, semantic, and ontological reasoning. Both anomaly detection module and reasoning engine module influence the attack adaptiveness (reflexivity) and self-healing of IAS, where decisions obtained through reasoning and learning are turned into actions. With this extensive cognitive computing modules, the final response from IAS to other interacting entities will be a secure and trusted one.

- Knowledge discovery module facilitates multi-faceted dimensions of advanced data analytics including regression analysis, supervised learning, unsupervised learning, and pattern-recognition. Discovered knowledge is shared with cognitive computing module for further learning. The proposed structure provides robust cyber resilience and autonomous operation of the system.

2. Background on

Cognitive Autonomy

Cognitive computing is a vital part of security in autonomous systems. In particular, malware and anomaly detection has become a biggest challenge with increase in sophistication in attacks such as file-less malware [1] and ransomware [2]. Behavior-based malware detection system (pBMDS) was proposed in [3]. The technique observes unique behaviors of applications as well as users and leverages Hidden Markov Model (HMM) to learn application and user behaviors based on two features: process state transitions and user operational patterns. One of the drawbacks of the HMM model is that it has very limited memory thus cannot be used for sequential data. In this project, we leverage hardware, software, and operating system data and apply long short-term memory units to identify anomalous behavior. We will also profile applications and malware using HW data (memory bytes and instruction sequences) to whitelist benign processes and blacklist malicious processes. In order to enable better results for LSTM deep learning methodologies, knowledge discovery and representation are important. We proposed a metadata labeling scheme, BFC, for information tagging and clustering by reversing the error correction coding technique known as Golay coding [4] [8]. The scheme utilizes 223 number of binary vectors of size 23 bits to profile features and cluster the data items. Since the method is built based on error correction scheme, it exhibits fault tolerance in wrongly labeled data. Similarly, we perform privacy-preserving knowledge discovery through perturbed aggregation in untrusted cloud [5]. In this project, we will use advanced data analytics to enable reasoning module for assisting attack adaptation and reflexivity of the system.

3. Cognitive Autonomy for Cybersecurity in Autonomous Systems

Decentralized machine learning is a promising emerging paradigm in view of global challenges of data ownership and privacy. We consider learning of linear classification and regression models, in the setting where the training data is decentralized over many user devices, and the learning algorithm must run on device, on an arbitrary communication network, without a central coordinator. We plan to utilize and advance COLA, a new decentralized training algorithm [23] with strong theoretical guarantees and superior practical performance. This framework overcomes many limitations of existing methods, and achieves communication efficiency, scalability, elasticity as well as resilience to changes in data and participating devices. We will consider fault tolerance to dropped and oscillation of nodes from connected to disconnected and attacks on the nodes. The learning has to be communication-efficient decentralized framework and free of parameter tuning. COLA offers full adaptively to heterogeneous distributed systems on arbitrary network topologies and is adaptive to changes in network size and data and offers fault tolerance and elasticity. IAS should have clear understanding of its operational context, it's won processes, and its interactions with neighboring entities. In this project, the cognitive computing module consists of three major components: (1) Malware / anomaly detection module, (2) Reasoning engine, and (4) Reflexivity engine. Cyber attribution data (system monitoring data or provenance data) is sent to cognitive computing engine for analysis where the system profiles the applications based on machine learning models. In this paper, we will focus on the cognitive autonomy property of the autonomous systems.

4. Malware and Anomalous Application Behavior Profiling with Deep Learning Model:

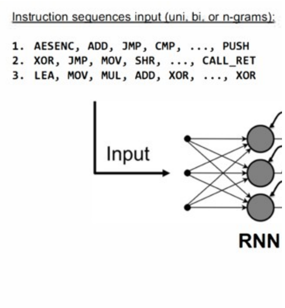


Figure 2: Recurrent Neural Network (RNN) model for application behavior profiling

We use instruction sequences executed in memory by application to understand the behavior of each application.
Input: n-gram sequences of instructions from memory
Output: Binary classification of benign or malicious

- Step 1: Define a finite set I of instructions {i_1, i_2, ..., i_n} in the system. Instructions are executed based on time epochs i.e., time-series data.
- Step 2: Given an observed sequence of {i_1, i_2, ..., i_n}, we find the set N of the top P sequences to be executed at time t. The size of the set N varies in each prediction and is determined by ngrams input as well as the clusters in the output of the model.
- Step 3: At time t, the sequence {i_1, i_2, ..., i_n} is benign if i_1 is in P, otherwise malicious.

Algorithm 1: Application Behavioral Profiling Algorithm

5. Malware and Anomaly Detection with Light-weight Machine Learning Models:

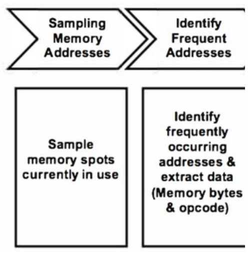


Figure 3: Malware/ anomaly Detection with Light-weight Machine Learning Methods

Advanced malware such as ransomware encrypts IAS data without authorization. Since it does not alter the system configurations and leave a footprint, it is difficult to detect them. But based on the executed instruction sequences and constants (also known as magic constants) used for encryption mechanism during malware execution, applications can be profiled. First, we will sample the address spots for every 1,000,000 instructions (fixed sampling). After a fixed period of time, we will calculate the frequently occurring addresses and their relevant process ids. A threshold T will be set for data extraction. For example, extract memory bytes and instructions from top T = 10% of the global list of sampled addresses (sorted in descending order based on their frequency of occurrence). Once opcode and memory bytes data is collected, we will extract features such as n-gram, bigram, unigram features, magic constants feature, cosine similarity with instructions occurrences, and standard deviation. Cosine similarity metric is one of the most efficient method to learn from large datasets [20]. It plays a crucial role in understanding similarity between two feature vectors when the magnitude of the vector is large or unspecified
i.e., it can either be unigram, bigram, or n-gram features. Given two feature vectors Vi = $\{f_{11}, f_{12}, ...\}$ and Vi = $\{f_{21}, f_{22}, ...\}$, where f_{11}, f_{21}, . . .are values of a particular feature, the cosine similarity is given as,

$$\text{Similarity}\,(V_1, V_2) = \frac{V_1 \cdot V_2}{||V_1||\,||V_2||} = \frac{\sum_{i=0}^{n} V_1^i\, V_2^i}{\sqrt{V1_i^2}\sqrt{V2_i^2}}$$

The cosine similarity lies

between O and 1. If the orientation of the two feature vectors is the same then the similarity between them is Cos O = 1 i.e., there is zero angle between them. But when the angle is 90° (the orientation of the feature vectors is at an angle of 90) then the similarity is Cos 90 = 0. The similarity score varies between [O, ½). Once the features are extracted, we will implement RF, SVM, and KNN learning models. K-NN is one of the simplest yet powerful classifier with high computational efficiency as well as accuracy [6].

6. Conclusion

We presented two approaches for detecting through profiling evasive malware applications. We use both light-weight machine learning models as well as deep learning models to profile and understand the behavior of autonomous systems. This multi-model approach is advantages when it comes to computational resources in mission critical systems. Based on the data and sample size, appropriate model can be selected for analysis. In particular, light-weight machine learning models use less computational resources and they have considerably less time complexity. On the other hand, LSTM model can provide robust classification with fundamental data, which enables IAS to understand evasive malware at basic level.

7. Acknowledgements

This research is funded by Northrop Grumman Corporation.

8. References

[1] Hopkins, Michael, and Ali Dehghantanha. "Exploit Kits: The production line of the Cybercrime economy?" In Information Security and Cyber Forensics (InfoSec), 2015 Second International Conference on, pp. 23-27. IEEE, 2015.

[2] [2] Kharraz, Amin, William Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. "Cutting the gordian knot: A look under the hood of ransomware attacks." In International Conference on Detection of Intrusions and Ma/ ware, and Vulnerability Assessment, pp. 3-24. Springer, Cham, 2015.
[3] Xie, Liang, Xinwen Zhang, Jean-Pierre Seifert, and Sencun Zhu. "pBMDS: a behavior-based malware detection system for cellphone devices." In Proceedings of the third A CM conference on Wireless network security, pp. 37-48. ACM, 2010.
[4] Mani, Ganapathy, Bharat Bhargava, and Jason Kobes. "Scalable Deep Learning Through Fuzzybased Clustering in Autonomous Systems." In IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AI.KE), pp. IEEE. 2018. http://www.cs.purdue.edu/homes/bb/aike2.pdf
[5] Mani, Ganapathy, Denis Ulybyshev, Bharat Bhargava, Jason Kobes, and Puneet Goyal. "Autonomous Aggregate Data Analytics in Untrusted Cloud." In IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AI.KE), pp. IEEE. 2018. http://www.cs.purdue.edu/homes/bb/aikel.pdf
[6] Prasath, V. B., Haneen Arafat Abu Alfeilat, Omar Lasassmeh, and Ahmad Hassanat. "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier-A Review." arXiv preprint arXiv:1708.04321 (2017).
[7] Bholowalia, Purnima, and Arvind Kumar. "EBK-means: A clustering technique based on elbow method and k-means in WSN." International Journal of Computer Applications 105, no. 9 (2014).

[8] Mani, Ganapathy, Nima Bari, Duoduo Liao, and Simon Berkovich. "Organization of knowledge extraction from big data systems." In *2014 Fifth International Conference on Computing for Geospatial Research and Application*, pp. 63-69. IEEE, 2014.