# Redefining Cybersecurity with AI and Machine Learning

Amol Dhondse
IBM Software Lab(ISL)
Bengaluru,
Indiaamol.dhondse@in.ibm.
com

Sachchidanand Singh
IBM Software Lab(ISL)
Pune, India
sachcsin@in.ibm.com

*Abstract*— **In the age of digital transformation with adoption of Cloud and mobile computing and ever-increasing Internet of Things(IoT) devices, the cybersecurity risks and threat levels are increasing at a rapid pace. The data is spread across systems, devices and cloud leading to growing attack surface and increased frequency of the security attacks. IoT is extended to drones, driver-less cars, industrial equipment, smart buildings, consumer goods, home appliances leaving us with more vulnerable attack points. Organizations needs to have effective information security management system (ISMS) in place to proactively detect, react to security threats with reduced time to discover any potential breach. This paper highlights how Artificial Intelligence(AI) and Machine Learning(ML) can redefine cybersecurity to detect, prevent organizations from security threats and data breaches.**

*Index Terms*— **Cybersecurity, Artificial Intelligence(AI), Machine Learning(ML), K Nearest Neighbors (KNN), Support Vector Machines (SVM), Markov Decision Process, Q-learning, Temporal Difference (TD), Attack Vector, Attack Surface, Naive Bayes Classifier, Logistic Regression, Neural Networks, Data Security, Decision Trees, Random Forest, Principal Component Analysis (PCA), Distributed Denial of Service (DDoS), TensorFlow, Torch, Caffe, DeepLearning.**

## I. Introduction

IT market is rapidly shifting towards cloud which raises new challenges for security and cloud security solutions are quickly evolving to meet these challenges. The increasing number of smart phone users and bring your own device (BYOD) trends are boosting growth of cyber security industry. Thegrowingneedtopreventmaliciousattacksinorganizations is anticipated to accelerate overall growth of the cyber security market infuture.

ArtificialIntelligence(AI)drivencyberattackscanlearnand get better as they evolve over time. For example, ransomware attacks using machine learning to know what information to hold hostage. Similarly, phishing scams have become more convincing using Artificial Intelligence(AI) to mimic the writing style of closefriends.

Machine Learning(ML) can be used to automatically monitor, discover, classify and protect sensitive information like intellectual property, financial data, source code, security keys etc. Machine Learning(ML) predictive models continuously build and analyze sources of sensitive information and typical suspicious access patterns likechanges to access control lists pointing to deliberate overexposure of the information. Therefore, ability to foresee future attacks and provide proactive counter measures is expected from cybersecurity teams.

## II. What is CyberSecurity

Thepreventativemethodsusedtoprotectsensitivedataand networks information from attack or unauthorized access is referred as Cybersecurity. The risk of cyber-attacks can be minimized by effective cyber security and organizations or individualscanbeprotectedfromtheunauthorizedexploitation of systems, networks and technologies. Cyber security covers applicationsecurity,informationsecurity,disasterrecoveryand networksecurity.

Therearesomanymachinelearningtechnologiesoutofthe box available which can be leverage open-source frameworks like TensorFlow [1], Torch or Caffe. Attackers and intruders could use artificial intelligence(AI) to make their exploits better and attacks moreintelligent.

The cybersecurity teams should work out potentialnetwork vulnerabilities through rigorous penetration testing using AI agents.Thesecurityexpertorwhitehathackercandeliberately breakintoprotectedorganizationnetworkstodiscoverpossible hacker’s entryroutes.

## III. What is MachineLearning

Machine learning provides computers and systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine Learning algorithms can be divided into three broad categories- Supervised learning, Unsupervised Learning and Reinforcement Learning.

### *A. Supervised Learning [2], [3], [4], [5],[6]*

Supervised learning algorithms build model relationships and dependencies between input features and target prediction output. It can predict output values for new data based on relationships it learned from the previous data sets. The computer is trained with labeled data. Some examples of supervised learning are regression, K nearest neighbors(KNN), naive bayes, decision trees, support vector machines (SVM), random forest, and neural networks etc.

### *B. Unsupervised Learning*[2]

Unsupervised learning is mainly used in the pattern detection and descriptive modeling. The model is trained on unlabeled data i.e. there are no output categories or labels here based on which the algorithm can try to build model relationships. Unsupervised learning algorithms work on input data to mine for rules, detect patterns, group the data points to derive meaningful insights.

Examples of unsupervised learning algorithm are apriori algorithm, k-means clustering and association rules etc.

### *C. Reinforcement Learning* [7],[8]

Reinforcement learning algorithm collect information from interaction with the environment to take actions that would maximize the reward and minimize the risk. Reinforcement learningalgorithmcontinuouslylearnsfromtheenvironmentin an iterative fashion until it explores full range of the possible states.

The machine and reinforcement learning algorithm (called the agent) determine the ideal behavior within a specific context automatically, to maximize performance. The agent learns its behavior from simple reward feedback called as reinforcement signal. Some examples of reinforcement learning are Markov decision process, Q-learning, temporal difference (TD), deep adversarial networks etc.

## IV. ATTACK VECTOR AND ATTACKSURFACE

An attack vector is a path by which a hacker gets access to acomputerornetworkservertocreatemaliciousoutcome.The attack vectors enable hackers to exploit system vulnerabilities by programming to weaken the system defenses. Attack vectors may include e-mail attachments, viruses, pop-up windows etc. Hackers are constantly updating attack vectors to gain unauthorized access to computers and servers, therefore a defense method which is effective today may not remain so in future.

The attack surface of a software environment is sum of all the attack vectors from where an unauthorized user can extract datafromanenvironment.Thebestapproachforbasicsecurity measure is to keep attack surface as small as possible. The are several approaches to reduce attack surface like reduce entry points available to untrusted users, turn off unnecessary functionality,removeservicesrequestedbyrelativelyfewusers etc.

## V. STAGES OF SECURITY BREACH

Cybersecurity teams uses machine learning and big datato create predictive analyses of how, and when cyberattacks could occur.

### *A. VulnerabilityDiscovery*

The attacker tries to find issues inside the system to break in.Thereareseveralwaystodiscovervulnerabilitieslikecheck for known issues by known payloads or generate newpayloads to discover new issue. The attacker tries to generate abnormal behavior like putting unusual data in the request fields tocause anunexpectedresponsefromthetargetservice.Theartificial intelligence trained models by already discovered payloads for existing vulnerabilities, can suggest new payloads to discover new issue with better probability.

### *B. Exploitation*

The attackers apply their knowledge and experience at the exploitation phase to gain access by using a previously discovered vulnerability. For well-known issues, this process can be automated by simply coding each step. The attacker needs to find the right way to penetrate a system environment or application infrastructure. Artificial intelligence can help to adapt an exploit for a given environment faster than human sinceitcangenerateexploitvariantsandrunthemmuchfaster.

### *C. Post-exploitation*

The attacker after exploiting the first issue would godeeper to discover new issue and exploit them. Any well-designed infrastructure consists of several isolated layers and attackers can easily replicate steps to exploit new layers after compromising onelayer.

### *D. Data Theft*

The attackers once gain access to different layers of the system start stealing vital customer data like credit card, passwords, social security number (SSN) etc. At times it may not be easy to steal a lot of data due to infrastructure restrictions like number of outbound filters. Artificial intelligence can help attackers to steal most valuable things first using data search and classification methods.

## VI. APPLYING MACHINE LEARNING TOSECURITY

The security use-cases can be broadly partition into two main group: first one where machine learning has made an impact and second one where machine learning is yet to make its marks and produce usable results.

First group comes under supervised learning category, here machine learning has made significant difference and all the problems having good labeled data for analysis. The second group falls into unsupervised category, we don't have labeled data and hence challenges in choosing the right approach.

The different machine learning problem types like dimensionality reduction is used to reduce number of dimensionsorfieldsofagivendata.Clusteringandassociation rules are used to create group of similar records and make it easier to analyze and understand large datasets. But these algorithms are of limited use when it comes to identifying attacksoranomalies.Howeverartificialintelligence(AI)canbe used to predict attacks by flagging any signs of potential security breach by analyzing network activity while actively comparing datasamples

## VII. MACHINE LEARNING PROBLEMTYPES

There are several types of machine learning problem types like classification, regression, clustering, dimensionality reduction etc.

### *A. Classification* [9], [10], [11], [12], [13], [14]

Classification is problem of identifying categories to which a new observation belongs based on the training data set containing observations whose category is already known. An example of classification problem could be to analyze a given image to determine type of object category it falls into or analyzingmedicaldatatodetermineifapersonisinahigh-risk group for a certain disease ornot.

Examples of algorithms used for supervised classifications problems are, Naive Bayes Classifier, Support Vector Machines, Logistic Regression, Neural Networks etc.

### *B. Regression*[5]

In regression we try to find relationship between variables in given data-set to make prediction on a continuous scale. For example,predictingthelikelihoodofrainonagivendatebased on historical data. The algorithms to solve regression problems are linear regression, non-linear regression, Bayesian linear regressionetc.

### *C. Clustering*[2],[20]

In clustering a given data-set is divided into many groupsof similar entities or data points. The purpose of clustering is tosegregate groups with similar traits and assign them intoclusters to further discover inherent groupings in the input data.Some of the popular clustering applications are anomalydetection, image segmentation, recommendation engines &market segmentation etc. The most widely used clustering algorithms are K-mean clustering and hierarchical clustering.

### *D. Dimensionality Reduction*[10]

In dimensionality reduction, an attempt to minimize number of dimensions for a given data-set is made so that information is presented concisely with no loss in the information. Dimensionality reduction technique is used to obtain better features for a classification or regression task in solving machine learning problems. Dimensionality reduction technique helps to reduce storage space, less dimensions leads to less computing, removes redundant features, allows to plot data and visualize precisely.

There are several methods to perform dimensionality reduction like Decision Trees, Random Forest, Principal Component Analysis (PCA) and Factor Analysis etc.

## VIII. Machine Learning Applied to SecurityProblems

### *A. DOS/DDOS Detection*[15]

Distributed Denial of Service(DDoS) is a type of DoS attack where multiple compromised systems, which are often infected with a Trojan, are used to target a single system. Victims of a DDoS attack consist of both the end targeted system and all systems maliciously used and controlled by the hacker in the distributed attack.

We can detect DDoS attacks using a supervised learning model by Support Vector Machines (SVM) [16], [19], which captures network traffic, filters HTTP headers, normalizes the data based on the operational variables: rate of false positives, rate of false negatives, rate of classification and then sends the information to correspond SVM's training and testing sets.

The detection of Distributed Denial of Service(DDoS) is a classification problem in context of machine learning.

### *B. Phishing Detection*[1]

Phishing is a form of fraud in which the attacker tries to steal sensitive information such as login credentials or account information by sending as a reputable entity or person in email or other communication channels.

This type of attack uses social engineering techniques to steal confidential information - the most common purpose of such attack targets victim's banking account credentials. Phishing attacks send spoofed emails to users which lead them to malware infected websites designed to appear as realon-line bankingwebsites.Emailsreceivedbyusersinmostcaseslooks authentic and contains direct request to verify some account information, credentials or credit card numbers by following the provided link. The request will be accompanied by a threat that the account may become disabled or suspended if the mentioned details are not being verified by theuser.

The detection of phishing domains is a classification problem, so it means we need labeled data which has samples as phishing domains and legitimate domains in the training phase. The dataset which will be used in the training phase is a very important point to build successful detection mechanism. We must use samples whose classes are precisely known. Otherwise, the system will not work correctly if we use samples that we are not sure about. We can use Decision Tree Algorithm for Phishing detection.

### *C. Detection of new classes of malware* [17], [18],[20]

A malware refers to malicious software perpetrators dispatched to infect individual computers or an entire organization's network. It exploits target system vulnerabilities, such as a bug in legitimate software like web application plugin which can be hijacked. A malware infiltration can be disastrous with consequences of data theftor the crippling of network systems. There are several malware types like Ransomware, Worms, Trojan, Spywareetc.

The hardware-assisted malware detection is based on monitoring and classifying memory access patterns using machine learning. This provides for increased automation and coverage through reducing user input on specific malware signatures. An online framework for detecting malware could uses machine learning to classify malicious behavior based on virtual memory access patterns. This type of approach comes under classification and clustering problems.

## IX. Summary andConclusion

We are moving towards a game of machine versusmachine and hence it becomes crucial to make sure AI applications learn to defend much faster than they learn toattack.

Machine learning algorithms will improve security solutions, helping human analysts triage threats and close vulnerabilities quicker. But they are also going to help threat actors launch bigger, more complex attacks. This has naturally led to fears that this is AI vs AI. Therefore, continuous innovation and industry collaboration will be critical for technology like artificial intelligence(AI), machine learning, predictive analytics to identify anomalous behavior, adapt to changing risk environment, and get ahead of the cyber criminals. Cybersecurity industry is dealing with a widening talent gap, and organizations and firms are hard-pressed to fill vacant security posts.

Machine Learning is cybersecurity's answer to detecting advanced breaches. Artificial intelligence(AI) can find innovativewaystofastdiscovervulnerabilities,identifycritical datapronetosecuritybreaches,unlikethepresent-daysituation when they are able to speed up a step-by-step attack scenario defined by humans. The cybersecurity analytics help to predict cyberattacks before they occur and AI techniques such as machine learning and deep learning can be used to find vulnerabilities that may be difficult for the security team to find.

## REFERENCES

[1] J. B. Fraley and J. Cannady, "The promise of machine learning in cybersecurity," SoutheastCon 2017, Charlotte, NC, 2017, pp. 1-6. doi:10.1109/SECON.2017.7925283

[2] Y. Xin et al., "Machine Learning and Deep Learning Methods for Cybersecurity," in IEEE Access. doi: 10.1109/ACCESS.2018.2836950

[3] I. Medeiros, N. Neves and M. Correia, "Detecting and Removing Web Application Vulnerabilities with Static Analysis and Data Mining," in IEEE Transactions on Reliability, vol. 65, no. 1, pp. 54-69, March 2016. doi:10.1109/TR.2015.2457411

[4] R. Komiya, I. Paik and M. Hisada, "Classification of malicious web code by machine learning," 2011 3rd International Conference on Awareness Science and Technology(iCAST), Dalian, 2011, pp. 406-411. doi: 10.1109/ICAwST.2011.6163109

[5] S. O. Uwagbole, W. J. Buchanan and L. Fan, "An applied pattern-driven corpus to predictive analytics in mitigating SQL injection attack," 2017 Seventh International Conference on Emerging Security Technologies (EST), Canterbury, 2017, pp. 12-17.doi:10.1109/EST.2017.8090392

[6] M. Stampar and K. Fertalj, "Artificial intelligence in network intrusion detection," 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2015, pp. 1318-1323. doi: 10.1109/MIPRO.2015.7160479

[7] Q. Chen and R. A. Bridges, "Automated Behavioral Analysis of Malware: A Case Study of WannaCry Ransomware," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, 2017, pp. 454-460. doi: 10.1109/ICMLA.2017.0-119

[8] D. C. Le, A. Nur Zincir-Heywood and M. I. Heywood, "Data analytics on network traffic flows for botnet behaviour detection," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-7. doi: 10.1109/SSCI.2016.7850078

[9] A. Joshi and V. Geetha, "SQL Injection detection using machine learning," 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kanyakumari, 2014, pp. 1111-1115. doi: 10.1109/ICCICCT.2014.6993127

[10] M. Ito and H. Iyatomi, "Web application firewall using character-level convolutional neural network," 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA), BatuFeringghi, 2018, pp. 103-106. doi: 10.1109/CSPA.2018.8368694

[11] P.Likarish, E. Jung and I. Jo, "Obfuscated malicious javascript detection using classification techniques," 2009 4th International Conference on Malicious and Unwanted Software (MALWARE), Montreal, QC, 2009, pp. 47-54. doi: 10.1109/MALWARE.2009.5403020

[12] R. Wang, X. Jia, Q. Li and S. Zhang, "Machine Learning Based Cross-Site Scripting Detection in Online Social Network," 2014 IEEE Intl Conf on High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS), Paris, 2014, pp. 823-826. doi:10.1109/HPCC.2014.137

[13] M. Kruczkowski and E. N. Szynkiewicz, "Support Vector Machine for Malware Analysis and Classification," 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies(IAT), Warsaw, 2014, pp. 415-420. doi: 10.1109/WI-IAT.2014.127

[14] S. Kumar, A. Viinikainenand T. Hamalainen, "Machine learning classification model for Network based Intrusion Detection System," 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, 2016, pp. 242-249. doi: 10.1109/ICITST.2016.7856705

[15] G. C. Y. Tsang, P. P. K. Chan, D. S. Yeung and E. C. C. Tsang, "Denial of service detection by support vector machines and radial-basis function neural network," Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), 2004, pp. 4263-4268 vol.7. doi: 10.1109/ICMLC.2004.1384587

[16] I. Paik, "Improved malicious code classification considering sequence by machine learning," The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014), JeJu Island, 2014, pp. 1-2. doi:10.1109/ISCE.2014.6884429

[17] S. Kilgallon, L. De La Rosa and J. Cavazos, "Improving the effectiveness and efficiency of dynamic malware analysis with machine learning," 2017 Resilience Week (RWS),Wilmington, DE, 2017, pp. 30-36. doi: 10.1109/RWEEK.2017.8088644

[18] I. Firdausi, C. lim, A. Erwin and A. S. Nugroho, "Analysis of Machine learning Techniques Used in Behavior-Based Malware Detection," 2010 Second International Conference on Advances in Computing, Control, and TelecommunicationTechnologies, Jakarta, 2010, pp. 201-203. doi: 10.1109/ACT.2010.33

[19] S. O. Uwagbole, W. J. Buchanan and L. Fan, "Applied Machine Learning predictive analytics to SQL Injection Attack detection and prevention," 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, 2017, pp. 1087-1090. doi:10.23919/INM.2017.7987433

[20] G. Yuan, B. Li, Y. Yao and S. Zhang, "A deep learning enabled subspace spectral ensemble clustering approach for web anomaly detection," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 3896-3903. doi:10.1109/IJCNN.2017.7966347