

# Using Machine Learning to Protect User Data in Cloud

Shahil Mohammed  
1225369223  
Group 1-11  
smoham68@asu.edu

**Abstract**— The primary motivation for this project is to comprehend the benefits and drawbacks of traditional methods of user data protection in the cloud. This project also aims to discover machine learning methods to help protect user data in the cloud while also reinforcing traditional data protection methods.

**Keywords**—SVM, DDoS, DoS, Man-in-the-middle, Phishing, Random Forest, Virtualization

## I. OVERVIEW

Today, cloud computing is one of the most important technologies we use because it powers the majority of the web applications we use. Individuals frequently lack the computational resources required to train a machine-learning model with massive amounts of data. Cloud computing comes to the rescue in such situations. Cloud providers such as Google Cloud, Amazon Web Services, Microsoft Azure, and others are raking in big bucks by offering cloud computing as a service. However, as more industries adopt cloud architecture on a daily basis, there is a growing demand for data stored in the cloud to be secure. There are numerous ways for a user's data to be stolen from the cloud. For example, the cloud machine could be compromised, the cloud software could be vulnerable to attacks, and there could be side-channel attacks on cloud services. Cloud computing has become increasingly popular among businesses over the last decade. As users, we rely on cloud-based software to safeguard sensitive information such as Social Security numbers, credit card information, home addresses, phone numbers, emails, and so on. As a result, we must be able to prevent this vital information from being accessed by unauthorized means such as DDoS, man-in-the-middle attacks, phishing attacks, zombie attacks, and others. While traditional information security methods may be effective, machine learning may allow us to achieve greater security and reinforce traditional methods. We are only concerned with application-level security and have complete faith in the cloud provider.

The main goal of this project's research is to identify various solutions discovered and implemented using machine learning to defend against various attacks that may compromise user data stored in the cloud. So,

throughout our project, we reviewed various papers implementing machine learning techniques in various aspects of cloud computing and user data protection. We gathered data from various papers on the accuracy of different machine learning algorithms in detecting different types of attacks and attempted to determine which machine learning algorithm is best for which scenario. While we were unable to examine all types of defense mechanisms and machine learning algorithms, we did our best to identify some of the most popular and recent ones. We did not investigate how attacks on cloud systems are carried out because we were primarily concerned with the defense mechanism. Furthermore, we are assuming that the cloud provider can be trusted. As a result, there are no discussions about hypervisor or operating system security. Side-channel attacks are one of the topics we haven't covered in depth. We have mainly focused on the papers which provide machine learning-based solutions to modern cloud computing-related problems and tried to find solutions which are viable and can be used in modern-day cloud infrastructure without significant performance hits.

### A. Different types of security attacks on the cloud

Cloud computing is widely regarded as a significant and beneficial shift in IT architecture, but additional security work is still required to mitigate this technology's shortcomings. Because a significant amount of information, both personal and business-related, is stored in the Cloud's data centers, any security flaws or vulnerabilities associated with the Cloud must be discovered and eliminated [1]. Attacks can come from well-known sources like Address Resolution Protocol (ARP), IP spoofing, Denial of Service (DoS), and so on, but they can also come from unknown sources like zero-day attacks. Machine learning (ML) techniques can be extremely useful in detecting and preventing cyberattacks of this type [5].

The attacks most often discussed in Cloud computing are the following [6]:

1. **Denial of Service (DoS) attack:** An attempt to prevent users from accessing desired resources is known as a denial of service (DoS) attack. To initiate a DDoS attack, many computers are utilized in concert.

2. **Zombie attack:** A zombie attack occurs when an attacker overwhelms a target with requests from seemingly unrelated network hosts. This type of attack causes Cloud to behave unexpectedly, which in turn affects the availability of Cloud services.
3. **Phishing attack:** The goal of a phishing assault is to deceive users into giving away their sensitive information by leading them to a bogus website. A malicious actor may set up shop in the cloud and utilize it to mask their own accounts and services, or the accounts and services of unsuspecting victims.
4. **Man-in-the Middle attack:** In the Cloud, an intruder can access information exchanged between data centers by performing a "Man in the Middle" attack, in which he or she compromises the communication link between two users [7].

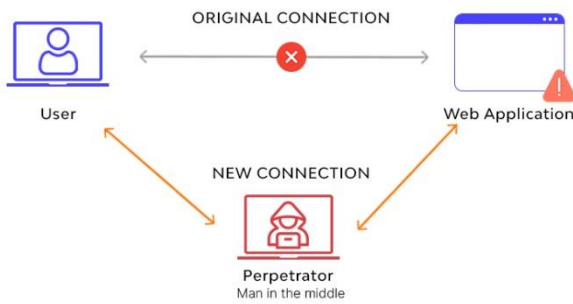


Figure 1: Man-in-the-middle attack

## B. Machine learning algorithms to detect threats

### Random Forest for IDS:

An intrusion detection system (IDS) is a piece of hardware or software that monitors a computer network or system for unethical behavior or policy violations. A random forest is a meta-estimator composed of several different decision trees. It is used to solve problems involving categorization and regression [3]. A random forest is a method for improving prediction accuracy by averaging the outcomes of several different decision trees.

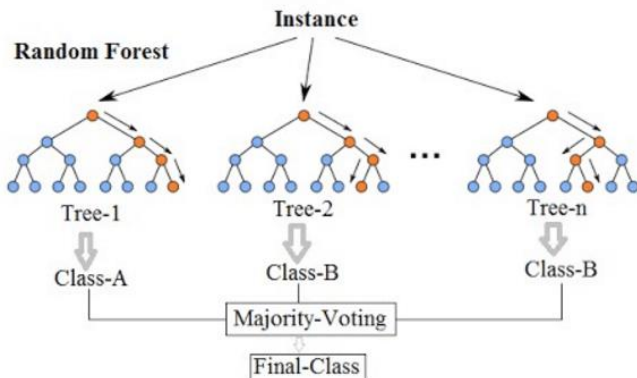


Figure 2: Random Forest classification

The random Forest method is a popular choice for a variety of machine learning tasks. It is simple to use and can frequently produce satisfactory results with little modification. This is a must-have advantage because we need the IDS to be dynamic in nature, and we need to be able to identify other patterns as quickly as possible with the least amount of interference. Allowing cloud providers to push hot updates in the shortest amount of time.

### Using MLP to detect phishing attacks in the cloud:

Phishing attacks are a type of cyber-attack in which people are tricked into providing personal information such as passwords or credit card numbers by using fake emails or websites. The attacker then uses this information to steal money or the victim's identity. Hackers are becoming more vigilant, most of these methods are becoming obsolete. Deep learning, a subdomain of AI with superior learning capabilities to traditional ML methods, can be used in this case. DL methods are computationally intensive models, but with the right training and data, cloud service providers can deploy them on NAT servers [2]. A multilayer perceptron (MLP) is a type of artificial neural network capable of learning complex data patterns. An MLP is made up of three layers: an input layer, hidden layers, and an output layer. To learn complex patterns in the data, the hidden layers employ a nonlinear activation function, such as a sigmoid or rectified linear unit (ReLU). To generate the final output, the output layer employs a linear activation function.

### Detecting DoS attacks using SVM based framework:

DoS attacks can occur when VMs consume all available physical resources, preventing the hypervisor from supporting any additional VMs. This attack is primarily the result of virtualization, which is the foundation of recent cloud computing architecture. Virtualization allows emulating a specific computer system and sharing physical resources. We propose a flexible detection framework based on the SVM learning technique to address the issues [4]. SVM is a classification technique that uses nonlinear mapping to convert original data into higher-dimensional data to find a hyperplane that best separates training tuples based on their classes. The following is a summary of our framework. The hypervisor gathers some features to train the SVM classifier to distinguish between normal VM activity and DoS attacks.

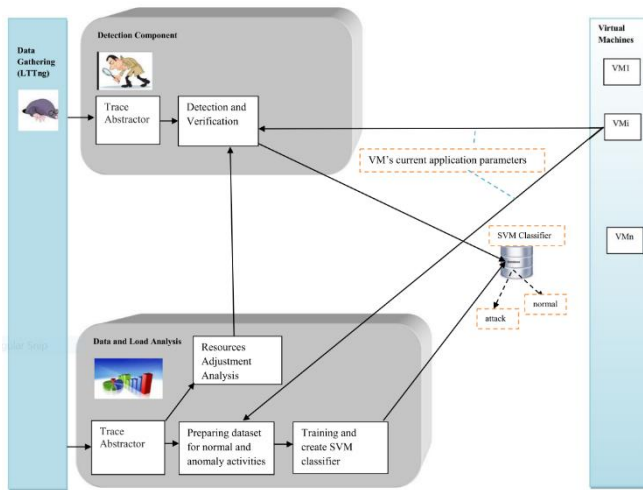


Figure 3: Architecture of SVM based framework

## II. CONTRIBUTION

- Identifying the most popular ML algorithm used for attack detection in the cloud is Support Vector Machine (SVM).
- Examining the prevalent security threats: distributed denial-of-service (DDoS) and data confidentiality
- Exploring different types of vulnerabilities in virtualization, internet protocols and management interface.
- Analysis of research papers using a methodology that involves search strategy, quality assessment rules and data extraction strategy.
- Identification of advantages and drawbacks of traditional machine learning algorithms used to secure data in the cloud.
- Critical analysis of performance and evaluation metrics used to classify the machine learning algorithms.
- Assisted the leader and deputy leader in writing weekly reports and setup Gantt charts.
- Worked with group to prepare the final project report and helped in proof-reading.
- Participated in developing presentation slides and helped resolve deficiencies.

## III. LESSONS LEARNED

- Understanding of various levels of risks and threats that are directly related to confidentiality, availability, and integrity of data.
- Knowledge of the lifecycle of the data and different stages the data undergoes in the cloud storage system.
- Understanding of possible phishing attacks in the cloud and ways to neutralize them using deep learning techniques such as ANN and MLP.
- Better understanding of usage of evaluation metrics to analyze and compare different ML models.

- Knowledge of different kinds of vulnerabilities exposing the cloud security.
- Ability to efficiently extract important information from a research paper after analyzing the content thoroughly.

## IV. REFERENCES

- [1] A. B. Nassif, M. A. Talib, Q. Nasir, H. Albadani and F. M. Dakalbab, "Machine Learning for Cloud Security: A Systematic Review," in IEEE Access, vol. 9, pp. 20717-20735, 2021, doi: 10.1109/ACCESS.2021.3054129.
- [2] I. Saha, D. Sarma, R. J. Chakma, M. N. Alam, A. Sultana and S. Hossain, "Phishing Attacks Detection using Deep Learning Approach," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 1180-1185, doi: 10.1109/ICSSIT48917.2020.9214132.
- [3] T. Saranya, S. Sridevi, C. Deisy, Tran Duc Chung, M.K.A.Ahamed Khan, Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review, Procedia Computer Science, Volume 171, 2020, Pages 1251-1260, ISSN 1877-0509.
- [4] Abusitta, A., Bellaiche, M. & Dagenais, M. An SVM-based framework for detecting DoS attacks in virtualized clouds under changing environment. J Cloud Comp 7, 9 (2018). <https://doi.org/10.1186/s13677-018-0109-4>
- [5] G. Nenvani and H. Gupta, "A survey on attack detection on cloud using supervised learning techniques," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), 2016, pp. 1-5, doi: 10.1109/CDAN.2016.7570872.
- [6] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, "A survey on security issues and solutions at different layers of cloud computing," J. Supercomput., vol. 63, pp. 561–592, Oct. 2013, doi: 10.1007/s11227-012-0831-5.
- [7] D. Chen and H. Zhao, "Data Security and Privacy Protection Issues in Cloud Computing," 2012 International Conference on Computer Science and Electronics Engineering, 2012, pp. 647-651, doi: 10.1109/ICCSEE.2012.193.