Using Machine Learning to Protect User Data in Cloud

CSE 543: Group 1-11

Venkata Sriram Medida

ASU ID: 1225364335

vmedida@asu.edu

OVERVIEW

The majority of web applications we use now are powered by cloud computing, making it one of the fundamental technologies we utilize every day. Cloud computing as a service is in high demand, and several different companies, like Google Cloud, Amazon Web Services, Microsoft Azure, and others, are profiting greatly from it. There are many ways in which the data of a user can be stolen from the cloud. As users, we have confidence in the cloud-based software to safeguard our private data, including Social Security Numbers, credit card information, phone numbers, and email addresses. Therefore, we must be able to prevent any unauthorized methods, such as DDoS, man-in-the-middle assaults, phishing attacks, zombie attacks, and others, from accessing this crucial information. Even though the information security provided by the traditional methods is good enough, using machine learning methods we can achieve much more end to end data privacy and protection.

The major objective of this project is to identify various defense strategies that have been developed and put into operation using machine learning to prevent various attacks from leaking user data stored in the cloud. We have collected data regarding the effectiveness of various machine learning algorithms detecting various types of attacks throughout the course of our project, reviewed various papers implementing machine learning techniques in various aspects of cloud computing, and tried to determine which machine learning algorithm is best for what kind of scenario. However, it was not possible to analyze all kinds of defense mechanisms and machine learning algorithms, we did our best to identify some of the most well-liked and contemporary ones. We primarily focused on papers that offered machine learning-based solutions to issues linked to modern-day cloud computing, and we looked for solutions that were workable and could be implemented in contemporary cloud architecture without suffering appreciable performance penalties.

ISSUES FACED BY MACHINE LEARNING MODELS:

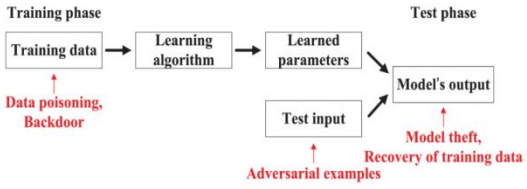
Despite the widespread use of machine learning-based applications, several security risks still exist for machine learning systems throughout their lifespans which makes it less likely to be reliable. The following categories describe the issues that machine learning models face [1] in relation to security:

1. Poisoning of the training set.

2. hidden passageways throughout the training set
3. Adversarial sample attacks.

4. piracy of models.

5.the recuperation of private training information.



From the above categories, models that use machine learning are more vulnerable to assaults that use adversarial examples than the other four models.

Adversarial example attacks are again divided into two: 1. Complete defense technique

2. Detection only technique.

For adversarial training, the "fast gradient sign technique" (FGSM) generation approach [1] is recommended since it increases the model's robustness. But using this approach always results in larger training data sets and higher training overhead. Many defensive models and detection approaches are introduced which are

used to mitigate these issues using the above stated defensive techniques.

The evaluation for the suggested approaches [1] can be achieved by following methods:

1. Design for security:

A framework is proposed for security evaluation of classifiers, and they simulate different levels of attacks by increasing the adversary's ability and knowledge.

2. Strong attacks: Perform security evaluation of machine learning algorithms by two aspects: (a) Under white-box attacks. (b) Under High Confidence attacks.

3. Evaluation Metrics: (a) It is suggested to take as many metrics as possible e.g., accuracy, precision etc. (b) Security evaluation curves can be used.

CLOUD COMPUTING AND APPROACH TECHNIQUES:

Data security and privacy protection are the main issues [2] that cloud computing users are concerned with. Cloud computing features include pervasive network access, on-demand self-service,

Quick resource elasticity, usage-based pricing, location-independent resource pooling, and risk transfer. Users' security concerns must be addressed in order to boost user trust in the cloud environment and encourage cloud computing adoption within enterprises. The cloud is always expanding because it can offer high performance computer services for less money.

The following approaches gives the techniques used [2] to reduce security threats in cloud:

1. RAID-like strategies and digital signature- Data Integrity can be obtained.
2. "Proofs of Retrievability" – Theoretical framework to realize remote data integrity checking by combining spot-checking and error correction code.
3. POR mechanism - to verify data storage across many clouds, establish redundancy between copies and implement availability and integrity checks.
4. Trusted platform module (TPM) – to check data integrity remotely.
5. Depot – can effectively resist attacks such as discarding and can also implement other safety Protections.
6. SPORC – uses a dependable cloud environment to create secure real-time communication and teamwork for multiple users.

VULNERABILITIES OBSERVED IN CLOUD AND TECHNIQUES TO RESOLVE:

The most frequent vulnerabilities observed in cloud computing environments are Denial of services (DOS), Domain Name Server Spoofing (DNS), Address Resolution protocol (ARP).

The proposed system (EIDC) [3] improves the assessment of node reliability when deciding the final packet classification. EIDC combines the output of the machine learning model with previously saved data (nodes attack types of categorization decision histories) for every packet received from any node in the network, and the most frequent choice is then made.

The accuracy of the model can be calculated using below formula:

Accuracy = TN + TP / number of packets

The simulation results demonstrate [3] that Enhanced Intrusion Detection and Classification (EIDC) can increase classification accuracy from 66% to as much as 90% and is an improvement over the conventional learning technique for attack detection. This demonstrates how the EICD outperforms the complicated tree at spotting traffic irregularities.

We further emphasize that the difference between complex trees and EICD is apparent from the first iteration forward and increases over time. The detection rate is roughly 90% for each node. Therefore, it is easier and quicker to identify and categorize rogue users the more packets and time iterations there are.

CLOUD COMPUTING ARCHITECTURE AND PRIVACY PROTECTION MODELS:

In general, privacy is defined as the collection, use, disclosure, storage, and destruction of personal data, regardless of country, culture, or jurisdiction (or personally identifiable information). The two primary concerns that are making consumers less reliant [4] on cloud computing are data security and privacy. Users should find out about a cloud provider's seven particular security measures, such as those pertaining to privileged user access, regulatory compliance, data placement, data separation, recovery, aid with investigations, and long-term viability, before making a commitment.

There are seven significant phases that make up the data life cycle: creation, transfer, usage, sharing, storage, archive, and destruction (Data life cycle refers to the entire process from data generation to data destruction).

Below figure gives the detailed structure of cloud architecture [4] that includes different layers of

security such as software security, platform security, infrastructure security and finally the auditing and compliance layer. The data passes through all these security layers and hence high levels of security standards are applied to it so that the data will be secure throughout its life cycle.

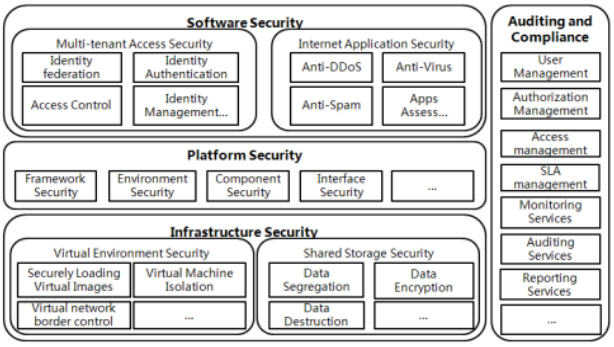


Figure 1. Cloud computing security architecture

CONTRIBUTIONS:

To analyze the different security related issues for machine learning models and their countermeasures.

- Finding out the methods to reduce the intensity of attacks in cloud environments thereby achieving Data Integrity and Security.
- Understanding the Various models proposed that can make machine learning models less vulnerable to attacks.
- Finding out the best evaluation technique among the proposed models can reduce the threats to a greater extent.
- Analysis of the simulation results of the EIDC algorithm to get a clear understanding of its benefits which makes it the best algorithm to be implemented.
- Done research on the data life cycle stages and the importance of data storage with respect to the aspects like integrity and confidentiality.

LESSONS LEARNED:

The following lessons were my takeaways from reviewing various reference materials and conducting my research for the project:

- Gained a complete understanding of the different security related issues that are being faced by Machine Learning and Cloud Environments.
- Acquired complete knowledge on the different security levels and their functionalities to secure data in the cloud environment.
- Learned about the various ML algorithms by comparing the efficiencies to take out the best that suits the given scenario.
- Got an understanding of the techniques and strategies that can be implemented to reduce the serious threats and vulnerabilities found in the cloud environment.
- Recognizing drawbacks and advantages of using standalone and hybrid machine learning models to prevent security attacks on the cloud.

REFERENCES:

[1] M. Xue, C. Yuan, H. Wu, Y. Zhang and W. Liu, "Machine Learning Security: Threats, Countermeasures, and Evaluations," in IEEE Access, vol. 8, pp. 74720-74742, 2020, doi: 10.1109/ACCESS.2020.2987435.

[2] Lakshmi K Ramachandrappa, Renukaradhya P C, Data Security Using Proxy based Methods in Cloud Computing," in International Journal of Advanced Scientific Innovation, Volume 02 Issue 03, August 2021.

[3] Z. Chkirbene, A. Erbad and R. Hamila, "A Combined Decision for Secure Cloud Computing Based on Machine Learning and Past Information," 2019 IEEE Wireless Communications and Networking Conference (WCNC), 2019, pp. 1-6, doi: 10.1109/WCNC.2019.8885566.

[4] 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.