Text

Description automatically generated

Cloud Security Using ML

Dissertation Proposal Research Methods

Date: 7-2-2022

Supervisor: Massoud Zolgharni Name: Abhay Jadav

ID: 21474963

Table of Contents

[Abstract 3](#_Toc94725709)

[Introduction 3](#_Toc94725710)

[Problem Statement 4](#_Toc94725711)

[Aim & Objectives 4](#_Toc94725712)

[Classification of Threats in Cloud Computing 5](#_Toc94725713)

[Model Analysis 6](#_Toc94725714)

[Machine Learning Algorithms 6](#_Toc94725715)

[Phases 1-2-3 7](#_Toc94725716)

[Literature Review 7](#_Toc94725717)

[Conclusion 8](#_Toc94725718)

# **Abstract**

Cloud computing is a rapidly expanding sector that is attracting the attention of large operating enterprises. It is necessary to identify and address several security risks that arise as a result of the advancement in this area of study. It is the purpose of this literature study to examine the dangers that may arise for an organization that is built on the cloud, and to consider the solutions that can be supplied to mitigate such threats. It is the goal of this study to identify as many probable threats as possible by training data and evaluating networks for recognizing odd patterns on networks for detecting threats with the assistance of machine learning and other technologies that are the most appropriate for the situation. The first stage in this study will be to develop a machine learning model that will be trained specifically for the purpose of identifying threats. I have even planned to update those network dangers that are out of the ordinary, in the hopes of avoiding the same threat from occurring on any other cloud network in the future. The following research publications have been selected as a starting point for our investigation: security and privacy problems in a cloud environment, as well as approaches to addressing these concerns in the cloud environment. In addition to this, I want to develop a machine learning model based on a regression model (which may be modified) for the purpose of applying my notions in this study.

# **Introduction**

In the previous literature review, we had talked in brief about the aims of the project and its working. Here in this assessment, the detailed information about the methodologies will be discussed along with the dataset which will be used in implementing our model. In the present era, there are many cloud computing companies which provides workspace for small and medium scale organizations. In concern for securing this environment, this assessment will discuss about different prediction algorithms and will choose the best fit algorithm. The dataset used in this research is downloaded from kaggel. There are many ongoing models which are being prepared. The aim for creating this model is to provide a secure environment in cloud computing. The first aim is training model by feeding features in different algorithms and checking the results. Along with targeting network logs for prediction, this model also aims for malware detection of ways for terminating it. The algorithms along with its advantages and disadvantages will be explained.

# **Problem Statement**

As most of the organizations are shifting towards cloud computing, there is a huge threat in cloud environment regarding its security. The threat towards data leak can result into huge loss for a organization. There are many such cases due to which some of the companies suffered huge loss and had to close. There are a variety of hazards that may damage a cloud network, with data breach being one of the most significant of these concerns. Data breaches in any corporation have the potential to bring the organization to its knees. In the first essay, the author discusses data breaches that have occurred in the real world and the consequences that have resulted for the company. Undoubtedly, one of the most intriguing cases is an Amazon employee who has been suspected of exposing the data of 106 million credit cards, as well as the majority of their personal information, to hackers. (Derrick Sampson, 2020) The unauthorized access to personal information may have major effects, and in certain cases, can even result in the demise of an organization. It is critical to identify and close any gaps that may exist as soon as they are discovered.

# **Aim & Objectives**

The aim of this research is to create machine learning model which can be deployed on cloud network as well as in cloud database. The network model will be made for a purpose of detecting unauthorized visitors in cloud network and the second aim is for creating a model which can be deployed in database for detecting malware and eliminating threats. The models will be created on basis of different datasets.

# **Classification of Threats in Cloud Computing**

When compared to the old system, the issues become more difficult to overcome as we go towards cloud computing. Three segments may be distinguished: the dynamic environment, the static environment, and the static environment. Because cloud computing is designed to operate in a dynamic environment, it might pose a security risk due to the fact that the majority of data will be kept online and will transit across a variety of different network services. In order to overcome these difficulties, it is necessary to test the system on a regular basis, which will aid us in identifying potential loopholes. (Deepak R Bharadwaj, 2018) The second issue that a cloud network has to deal with is the spread of its servers. Because it is dispersed over several sites, it makes it difficult to manage the data contained inside them. The third point of concern is the physical security of the facility. Cloud computing allows organizations to store all of their data in a distant place, which must be appropriately protected. If an intruder gains access to such systems, it is possible that data breaches may occur. Account hijacking, data breaches, malevolent insiders, and inadequate due diligence are all possible outcomes of poor access control. (Deepak R Bharadwaj, 2018) It is possible to calculate the number of workload hazards based on the quantity of work that is being performed in a cloud environment. A poorly constructed cloud structure may result in an overload on servers, which can end in the failure of the servers themselves. Many times, such crashes are caused by the introduction of an untrusted API on a server, which is why it is critical to maintain the network access limited. (Deepak R Bharadwaj, 2018).

# **Model Analysis**

The first step in creating model will be analysis and on basis of the analysis, different results can be obtained such as knowing the flow of traffic on a network as well as the features which can be useful in recognizing trusted and untrusted network. Analysis will be of two different types which will be performed on different datasets which are, the network traffic analysis. In network traffic analysis, the logs of network will be recorded and on basis of the logs, certain criteria will be defined and these categories will help us in differentiating the trusted and known users with the unauthorized users. The second analysis that will be used is in cloud organization. In a private cloud network, we will be using different malware detection process such as static analysis, dynamic analysis and many more. (Sunita Choudhary, 2020) With the help of these analysis, we will be using different methodologies such as ML prediction algorithms.

# **Machine Learning Algorithms**

In Machine Learning, there are many prediction models such as SVM, KNN, Native Based, Multilayer prescription, Linear Regressor and many more. (Sunita Choudhary, 2020) For predictions, algorithms such as random forest, decision tree and linear regressor can be used but the main disadvantage for these algorithms is they cannot be much effective on large scale data because of which it wont be suitable for cloud network. (Vrushang Patel, 2020)The best suitable algorithm for this model is SVM, KNN and XGBoost. These algorithms will be performed and on basis of results, the final algorithm will be selected. SVM will help us in classifying two different classes where as XGBoost can give quick results and will be based on decision tree concept. KNN can be considered as one of the best algorithms as it can be used for classification as well as regression methods. The only disadvantage of KNN is it can be slow when the number of features are increased. (Harrison, 2018)

# **Phases 1-2-3**

The first phase for completing this project is looking for appropriate dataset. For the development, I have chosen few different datasets but it can be updated in future. As soon as the dataset is finalized, the study of features will be done through which we will generate an idea for the dataset will be preprocessed accordingly. Data preprocessing will be the second phase. In this phase, the dataset will be molded in a way through which the prediction will be performed. As soon as the data is processed, data analysis will be done on dataset and necessary information will be noted. The last process in phase 2 will be of creating a final dataset on which the predictions will be performed. This dataset will be cleaned and designed in a way for achieving best prediction. In phase 3, the deployment of ML model will be done. For deploying ML algorithm, we will be implementing the final dataset into few different ML algorithms and based on the accuracy, the model will be finalized. Phase 4 will include the implementation and testing of model.

# **Literature Review**

In this literature review, we will be discussing about continuous data feeding and adaptive learning for network security, The ways of implementing ML model in clous, Framework for analyzing unusual activities in cloud, and predicting upcoming attacks. The data feeding can be done on a constant base with the help of ADWIN. (Pavol Mulinka, 2020) As of until now, the models we discussed had aim to detect on going treat and unwanted malware. The further aim for this research is for creating model which can help in predicting upcoming threats. (ELSAYED, 2019)

# **Conclusion**

This dissertation proposal has helped in understanding the need of securing cloud computing an the different ways through which the security can be provided. We have also discussed different algorithms which and the difference between those algorithms. This proposal has led to learning for possible outcomes in the field of cloud computing. There are many more uses of machine learning along with securing cloud network. It can also be used in making predictions for any organizations predicting the next target or by increasing the sales for the company.

# Bibliography

Deepak R Bharadwaj, A. B. M. C., 2018. *Cloud Threat Defense – a Threat Protection and,* Banglore: IEEE Xplore.

Derrick Sampson, M. M. C., 2020. *The Growing Security Concerns of Cloud Computing,* PA, USA: IEEE Xplore.

ELSAYED, M. A., 2019. *PredictDeep: Security Analytics as a Service for,* Canada: IEEE Xplore.

Harrison, O., 2018. *Machine Learning Basics with the K-Nearest Neighbors Algorithm,* s.l.: Towards Data Science.

Pavol Mulinka, P. C. J. V., 2020. *Continuous and Adaptive Learning over Big,* Prague: IEEE Xplore.

Sunita Choudhary, A. S., 2020. *Malware Detection & Classification using Machine,* Lakshmangarh: IEEE Xplore.

Vrushang Patel, S. C. a. T. H., 2020. *Predicting Future Malware Attacks on Cloud,* Monitoba, Canada: IEEE Xplore .