# BBB

***koeradoera CCC***

# BBB

*Thesis submitted to the*
*Indian Institute of Information Technology Guwahati*
*for award of the degree*

*of*

**Master of Technology**

*by*

**koeradoera CCC**

**under the supervision of**

**Dr. AAAAA BBB**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**INDIAN INSTITUTE OF INFORMATION TECHNOLOGY GUWAHATI**

**May 2020**

## CERTIFICATE

*This is to certify that the thesis entitled* **“BBB”**, *submitted by* **koeradoera CCC** *to the somehwere, for the award of the degree of Master of Technology, is a record of bona fide research work carried out by him under my supervision and guidance. The thesis, in my opinion, is worthy of consideration for the award of the degree of Master of Technology in accordance with the regulations of the Institute. To the best of my/our knowledge, the results embodied in the thesis have not been submitted to any other university or institute for the award of any other degree or diploma*

Dr. AAAAA BBB,

Assistant Professor,

Department of Computer Science and Engineering

somehwere

Date:

## DECLARATION

I certify that

a. The work contained in this thesis is original and has been done by me under the general supervision of my supervisor(s).

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in writing the thesis.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

koeradoera CCC

# ACKNOWLEDGMENTS

I would like to thank my supervisor Dr Angshuman Jana for helping me to carry out this thesis under his guidance. I am also thankful to our Head of Department Dr. Rakesh Matam Sir for his valuable guidance from time to time. His guidance helped me in researching and writing the thesis.Also I would like to thank our panel members Dr. Srinibas Swain, Dr. Rohit Tripathi,Dr. Soumi Chattopadhyay and Professor Malay Ananada Dutta from the panel.

## ABSTRACT

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

# List of Abbreviations

| | |
|---|---|
| SaaS | Software as a Service |
| S3 | Amazon Storage Service |
| SOA | Service Oriented Architecture. |
| SWS | Simple Workflow Service |
| SLA | Service Level Agreement |
| VM | Virtual Machine |
| VPC | Virtual Private Cloud |

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

## 1.1 Motivation

Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing

## 1.2 Objective of the thesis

Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random
thing Some random thing Some random thing Some random thing Some random thing Some
random thing Some random thing Some random thing Some random thing Some random thing
Some random thing Some random thing Some random thing Some random thing Some random

thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

## 1.3 Contribution of the thesis

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 1.3.1 First contribution

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 1.3.2 Second contribution

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

## 1.4 Organization of the thesis

The rest of the thesis is organized as follows.
In **Chapter 2**,Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some

random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing
In **Chapter 3** Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing . In **Chapter 4** Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing . Finally, we conclude in **Chapter 5** Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing .

# Chapter 2

# Literature Review

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

## 2.1 Deployment Models of Clouds

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing .

**Private Cloud**: Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

**Public Cloud**: Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

**Community Cloud**: Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing .

**Hybrid Cloud**:Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 2.1.1 Cloud Service Categories

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing -

#### 2.1.1.1 IaaS

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

#### 2.1.1.2 PaaS

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

#### 2.1.1.3 SaaS

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 2.1.2 Security Concerns

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Security Issues with respect to Cloud Computing [**?**] Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing :

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some

random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing [**?**] to launch such attacks. Hence from security perspective it is important to identify such an issue that may exist in underlying cloud infrastructure.

In [**?**] Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

In [**?**] Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

In [**?**] Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing [**?**]. Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing In September

2016 [?] Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

## 2.2 Cloud Services Comparison

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some

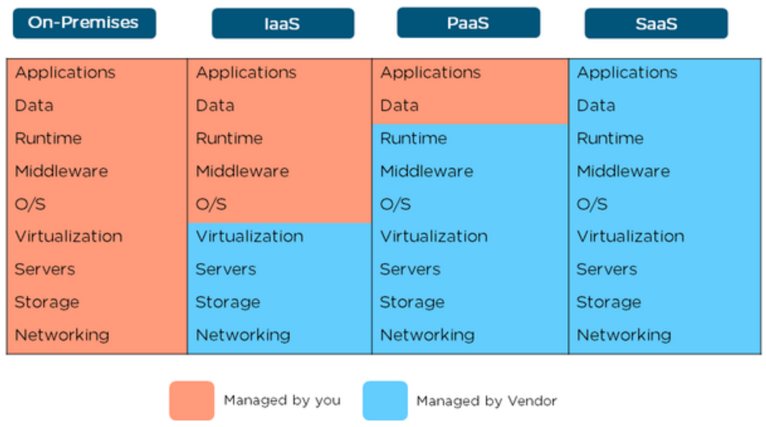


**Figure 2.1:** Cloud Services Comparison [?]

random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

## 2.3 Type of Anomalies

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

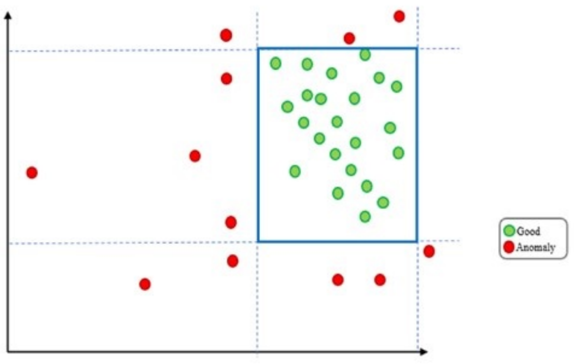


**Figure 2.2:** Simple Anomaly [?]

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing [?]

### 2.3.1 Global anomaly

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 2.3.2 Contextual Anomaly

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some

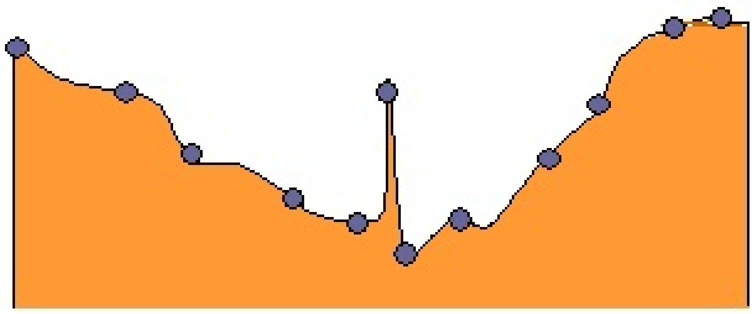


**Figure 2.3:** Global Anomaly [?]

random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

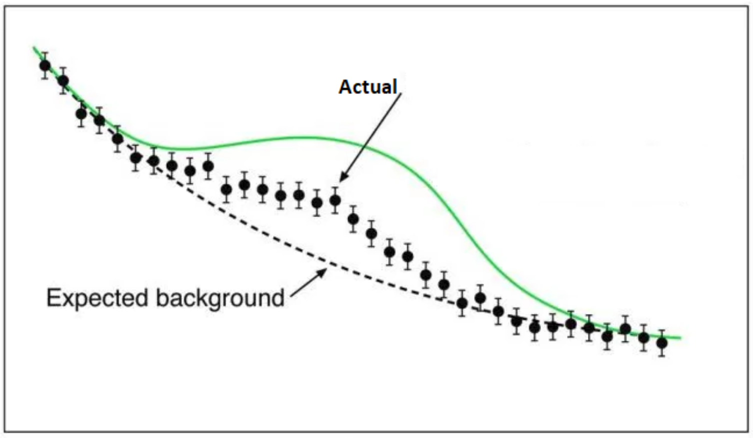


**Figure 2.4:** Contextual Anomaly Example [?]

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 2.3.3 Collective Anomaly

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

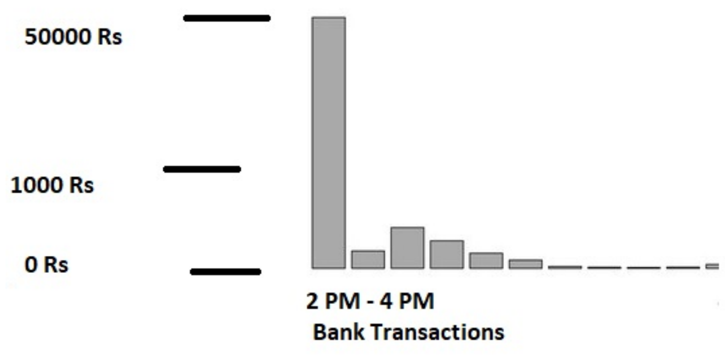


**Figure 2.5:** Collective Anomaly [?]

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

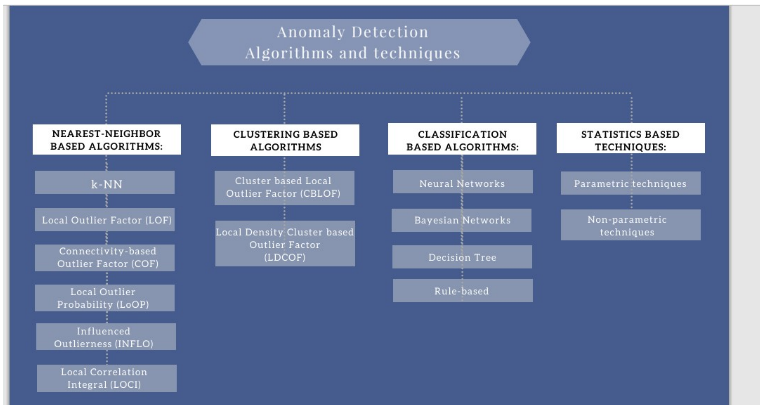

Anomaly Detection
Algorithms and techniques
NEAREST-NEIGHBOR BASED ALGORITHMS:
k-NN
Local Outlier Factor (LOF)
Connectivity-based Outlier Factor (COF)
Local Outlier Probability (LoOP)
Influenced Outlierness (INFLO)
Local Correlation Integral (LOCI)
CLUSTERING BASED ALGORITHMS
Cluster based Local Outlier Factor (CBLOF)
Local Density Cluster based Outlier Factor (LDCOF)
CLASSIFICATION BASED ALGORITHMS:
Neural Networks
Bayesian Networks
Decision Tree
Rule-based
STATISTICS BASED TECHNIQUES:
Parametric techniques
Non-parametric techniques

**Figure 2.6:** Anomaly Detection Algorithms [?]

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 2.3.4 Density-Based Anomaly Detection

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing thing Some random thing Some random thing Some random thing Some random thing

### 2.3.5 Clustering-Based Anomaly Detection

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing thing Some random thing Some random thing Some random thing Some random thing

### 2.3.6 Support Vector Machine-Based Anomaly Detection

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing thing Some random thing Some random thing Some random thing Some random thing Some

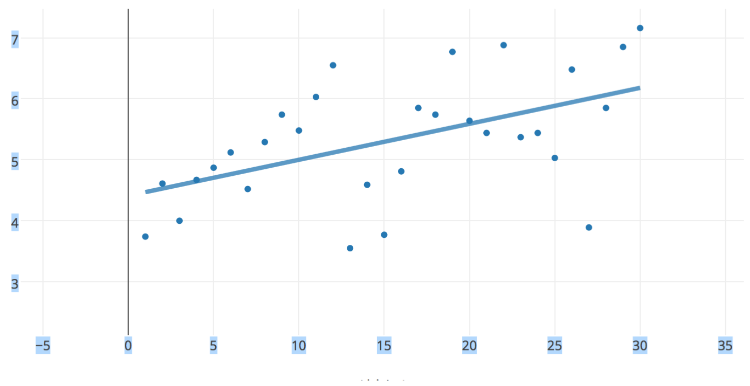


**Figure 2.7:** Anomaly detection using two variables [**?**]

random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

### 2.3.7 Anomaly detection algorithms categories

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

#### 2.3.7.1 K-nearest neighbor: k-NN

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

#### 2.3.7.2 Local Outlier Factor (LOF)

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

#### 2.3.7.3 K-means

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some

random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing
The user has to define the number of clusters in the early beginning.Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

#### 2.3.7.4 Support Vector Machine (SVM)

ASome random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

#### 2.3.7.5 Neural Networks Based Anomaly Detection

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

**Table 2.1:** Anomaly Detection Algorithms comparison [?]

| **Algorithm** | **Pros** | **Cons** |
|---|---|---|
| • K Nearest Neighbour<br>• K-NN | 1. Very easy to understand<br>2. Good for creating models that include non standard data types such as text | Large Storage requirements Computationally Expensive Sensitive to the choice of the similarity function for comparing instances |
| Local Outlier Factor(LOF) | Well-known and good algorithm for local anomaly detection | Only relies on its direct neighborhood .<br>Perform poorly on data sets with global anomalies. |
| K Means | Low Complexity<br>Very easy to implement | Each cluster has pretty equal number of observations<br>Necessity of specifying K<br>Only work with numerical data |
| Support Vector Machine (SVM) | Find the best separation hyper-plane.Deal with very high dimensional data.<br>Can learn very elaborate concepts. Work very well | Require both positive and negative examples. Require lots of memory.<br>Some numerical stability problems.Need to select a good kernel function |
| Neural networks based anomaly detection | Learns and does not need to be reprogrammed.<br>Can be implemented in any application | Needs training to operate<br>Requires high processing time for large neural networks<br>The architecture needs to be emulated |

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing

# Chapter 3

# First Contribution

Movement of services is essential for IT employees and other employees it increases the flexibility.Ever increasing modernization increases challenges for IT professionals in terms of new demands and complexities for keeping the organization secure. It is difficult to get the full benefit of cloud apps and services, an IT team must find the right combination of supporting technologies while maintaining control to protect critical data.Day by day with increasing challenges keeping the systems secure and marinating their integrity has been a challenging task. This paper studies various anomaly detection techniques used by cloud service providers and various approaches which exist in literature and discuses the anomaly detection techniques using KDD data set.

**Keywords:** Cloud, Virtualization, Time Series Analysis,Anomaly Detection,Distance Based Clustering

## 3.1 Outlier Detection

Any deviation from normal is considered an anomaly.Outlier detection (also known as anomaly detection) is the process of finding data objects with behaviors that are very different from expectation. Such objects are called outliers or anomalies. The most interesting objects are those, that deviates significantly from the normal object. Outliers are not being generated by the same mechanism as rest of the data. [**?**]

## 3.2 Security Framework

In Cloud Computing services security is a series of documented processes and frame works.Which revolves around defining policies and procedures around the implementation. For any successful security implementation of security frame work we need to define policies and procedures with respect to cloud.The main point of having a frame work is to reduce the risk levels and organizations exposure to vulnerabilities.

**Control the use of apps**: First we identify the cloud apps,various cloud apps can be using infrastructure as a service, or platform as a service services used by your organization.One needs to study usage patterns, identify the risk levels and business readiness of apps against risks. Start managing them to ensure security and compliance. [**?**] To protect your sensitive information anywhere in the cloud: understand, classify, and protect the exposure of sensitive information at rest. Apply out-of-the box policies and automated processes to apply controls in real-time across all your cloud apps.Then protect against cyber threats and anomalies: Detect unusual usage patterns analyzed by behavior of app usage across cloud to identify ransomware,currency mining, compromised users or fake applications, analyze high-risk usage and remediate automatically to limit the risk to your organization. In this process you need to assess the compliance of your cloud apps: Assess if your cloud apps meet relevant compliance requirements including regulatory bodies and industry standards. Prevent data leaks to non-compliant apps, and limit access to regulated data.When an organization elects to store data or host applications on the public cloud, it loses its ability to have physical access to the servers hosting its information. As a result, potentially sensitive data is at risk from insider attacks.The extensive use of virtualization in implementing cloud infrastructure brings unique security concerns for customers or tenants of a public cloud service. Virtualization affects the relationship between the operating system and underlying hardware – be it networking,storage or even computing. This introduces an additional layer – virtualization – that itself must be properly configured, managed and secured. Security in Microsoft Systems is done based via policies in accounts.To create policies go to control inside it go to templates.Select a policy template from the list, and then choose (+) Create policy. Customize the policy (select filters, actions, and other settings), and then choose Create. On the Policies tab, choose the policy to see the relevant matches (activities, files, alerts). Tip: To cover all your cloud environment security scenarios, create a policy for each risk category. How can policies help your organization? You can use policies to help you monitor trends, see security threats, and generate customized reports and alerts. With policies, you can create governance actions, and set data loss prevention and file-sharing controls.

## 3.3 How Microsoft Detects Anomaly in Azure

Microsoft uses a technology called security center [**?**] in its cloud environments.To detect threats and reduce false positives Security Center collects data from Azure resources and network and

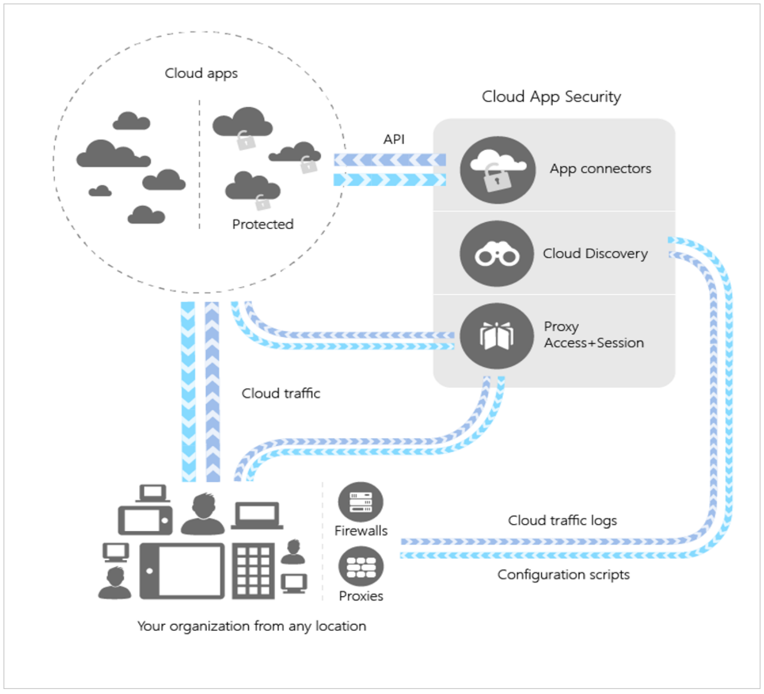


**Figure 3.1:** Security architecture [?]

then applies a lot of machine learning and big data algorithms to detect threat.These techniques as more advance than normal signature based anomaly detection techniques.It is impossible to manually identify the attack and predict the attack when it might happen.So the Microsoft Security Center uses following technologies

1. Intelligent Threat Intelligence
2. Behavioral Analytics
3. Fusion Analytics

### 3.3.1 Intelligent Threat Intelligence

It looks for bad actors by using the information obtained via Microsoft Products and services,Microsoft Digital Crimes Unit and Microsoft Security Response Centre along with external feeds.Microsoft has an immense amount of global threat intelligence. Telemetry flows in from multiple sources, such as Azure, Office 365, Microsoft CRM online, Microsoft Dynamics AX, outlook.com, MSN.com, the Microsoft Digital Crimes Unit (DCU), and Microsoft Security Response Center (MSRC). Researchers also receive threat intelligence information that is shared among major cloud service providers and feeds from other third parties. Azure Security Center can use this information to alert you to threats from known bad actors.

### 3.3.2 Behavioral Analytics

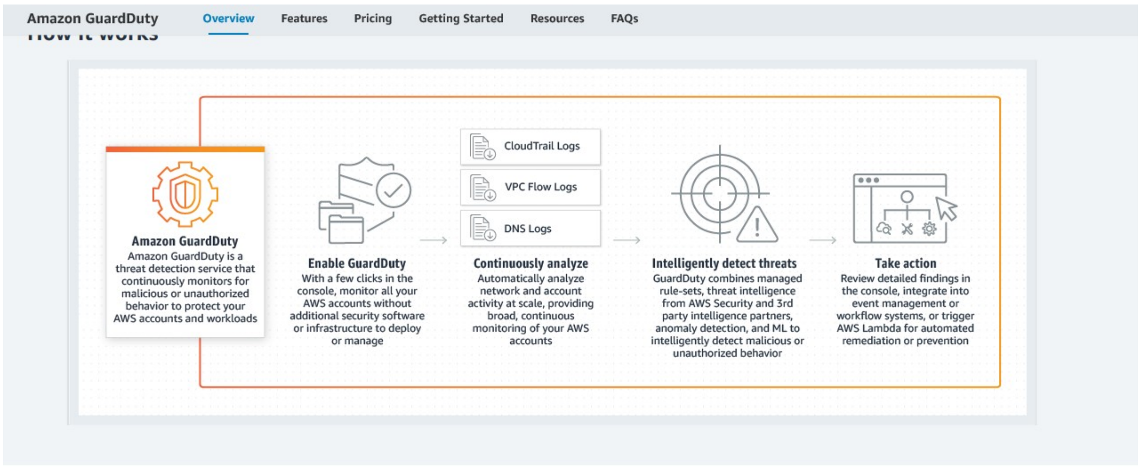
Behavioral analytics is a technique that analyses and compares data to a collection of known patterns. However, these patterns are not simple signatures. They are determined through complex machine learning algorithms that are applied to massive datasets. They are also determined through careful analysis of malicious behaviors by expert analysts. Azure Security Center can use behavioral analytic s to identify compromised resources based on analysis of virtual machine logs, virtual network device logs, fabric logs, crash dumps, and other sources. In addition, there's correlation with other signals to check for supporting evidence of a widespread campaign. This correlation helps to identify events that are consistent with established indicators of compromise.

### 3.3.3 Fusion Analytics

It uses statistical techniques to build a historical data which is based on usage patterns and any deviation from normal alerts those deviations [**?**] and it creates a baseline which if confirms to a potential attack vector then the particular usage is detected as an anomaly and in turn thus could be a security event.
Security Centre in Azure also works with connected partner solutions, like firewall and endpoint protection solutions. Microsoft uses **Fusion Analytics** [**?**] as the backbone of Security Centre's anomaly detection system.Fusion works by looking at various kind of alerts generated in Microsoft Azure ecosystem and then it tries to find pattern which could reveal attack progression indicating what should be next course of action.

## 3.4 How Amazon Finds Anomalies in AWS

Amazon uses a technology called guard duty [**?**].This service continuously monitors for threat detection service and continuously monitors bad behavior and protects aws accounts and workloads.In the cloud related services collection and aggregation of network related information and activities is simplified, but it is time consuming for security teams to continuously analyses event log data for potential threats. With the help of Guard Duty now you have an cost effective and intelligent solution which is used for threat detection in AWS cloud. Guard Duty uses machine learning, anomaly detection, and integrated threat intelligence to identify and prioritize potential threats. Guard Duty analyses tens of billions of events across multiple AWS data sources, such as AWS CloudTrail, Amazon VPC Flow Logs, and DNS logs. With a few clicks in the AWS Management Console, Guard Duty can be enabled with no software or hardware to deploy or maintain. By integrating with AWS Cloud Watch Events [?], Guard Duty alerts are actionable, easy to aggregate across multiple accounts, and straightforward to push into existing event management and workflow systems.



**Amazon GuardDuty**
Amazon GuardDuty is a threat detection service that continuously monitors for malicious or unauthorized behavior to protect your AWS accounts and workloads

**Enable GuardDuty**
With a few clicks in the console, monitor all your AWS accounts without additional security software or infrastructure to deploy or manage

**Continuously analyze**
Automatically analyze network and account activity at scale, providing broad, continuous monitoring of your AWS accounts

**Intelligently detect threats**
GuardDuty combines managed rule-sets, threat intelligence from AWS Security and 3rd party intelligence partners, anomaly detection, and ML to intelligently detect malicious or unauthorized behavior

**Take action**
Review detailed findings in the console, integrate into event management or workflow systems, or trigger AWS Lambda for automated remediation or prevention

**Figure 3.2:** Amazon Guard Duty [?]

## 3.5 How Google finds anomalies in GCP cloud

Google uses an open source library Forseti [?] which can

1. Detect unusual firewall behaviors between snapshots.
2. Alert users to any unusual behaviors and provide a comparison with expected behaviors.
3. Provide potential remediation steps.

The key elements for this technology are firewall rules.Firewall rules can be inbound or outbound. A firewall rule can either allow traffic based on IP address or ports.Firewall rules are applied to those instances which are associated with the user who has created his cloud in GCP setup. The figure below shows the technical architecture as how this has been implemented in GCP

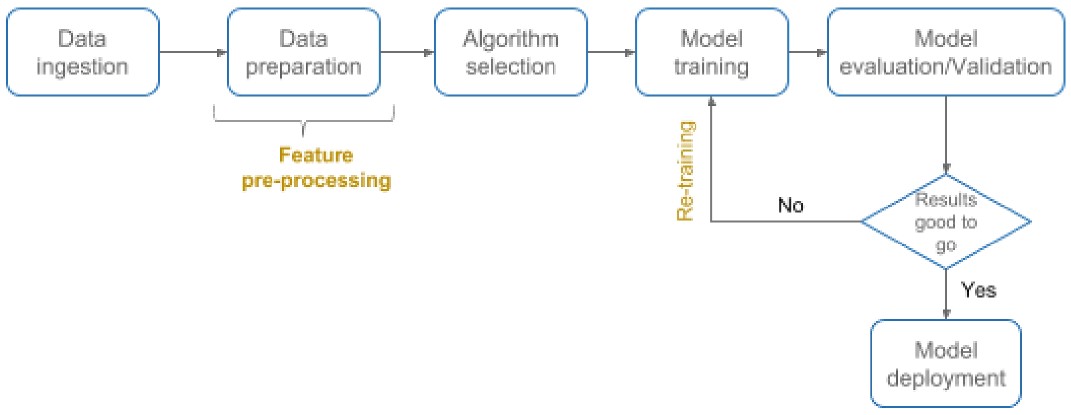


**Figure 3.3:** GCP Anomaly Detection [?]

## 3.6 Finding Anomaly in Cloud

The main goal of an anomaly detection system is to discriminate the occurrence of hostile activities from the normal events, and such analysis must be accomplished in a sufficiently flexible and effective way to keep up with the continuously evolving world of cyber security where new, previously unknown, anomalies can continuously emerge over time. In doing this, it must either try to model any kind of attack or anomalous event that can affect the network (there are thousands of known ones) or simply construct a sufficiently general model describing the normal traffic.Such model is usually built on the basis of training data, and used in classifying previously unseen or suspicious events. Classification is the fundamental task in unattended detection, by which the system “learns” to automatically recognize complex traffic patterns, to distinguish between different events based on the corresponding patterns, and to make “intelligent” decisions. Specific machine learning techniques, such as Neural Networks or Support Vector Machines are used in an anomaly detection system.

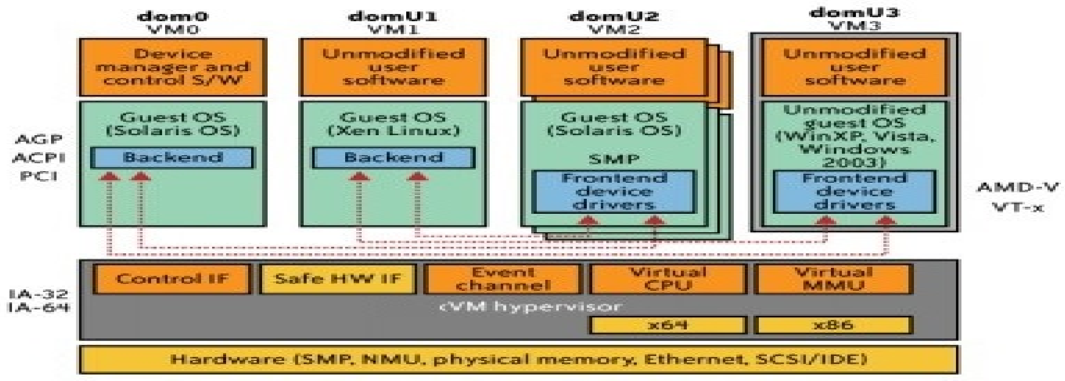


**Figure 3.4:** Hypervisor working [?]

### 3.6.1 Building the Anomaly Detection System

An Anomaly Detection System can be build using a generalization capability from training data which is needed to correctly classify future data as normal or abnormal. These resulting approaches can be categorized as generative or discriminating. A generative approach builds a model solely based on normal training examples and evaluates each testing case to see how well it fits the model. A discriminating approach, on the other hand, attempts to learn the distinction between the normal and abnormal classes. Thus, based on the characteristics of training data used to build the model, anomaly detection can be divided into three broad classes which appear below the table (See Table 3.1).

| Technique | Category |
|---|---|
| Supervised anomaly detection | A |
| Semi Supervised Anomaly Detection | B |
| Unsupervised anomaly detection | C |

**Table 3.1:** Type of anomaly detection techniques.

## 3.7 Detecting the deviation

In particular, in the context of abuse and network intrusion detection [?], the interesting objects are often not rare objects, but unexpected bursts in activity. This pattern does not adhere to the common statistical definition of an outlier as a rare object, and many outlier detection methods (in particular unsupervised methods) will fail on such data, unless it has been aggregated appropriately. Instead, a cluster analysis algorithm may be able to detect the micro clusters formed by these patterns.Three broad categories of anomaly detection techniques exist.i.e. Statistical Anomaly Detection Systems, Data Mining Based Anomaly Detection Systems , Machine Learning Based Anomaly Detection Systems. [**?**] Unsupervised anomaly detection techniques detect anomalies in an un labeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. Supervised anomaly detection techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier (the key difference to many other statistical classification problems is the inherent unbalanced nature of outlier detection). Semi-supervised anomaly detection techniques construct a model representing normal behavior from a given normal training data set, and then test the likelihood of a test instance to be generated by the learned model.

Anomaly detection [**?**] finds its use in a lot of fields such as fraud detection, fault detection, health monitoring, intrusion detection,event detection in sensor networks, finding disturbances in ecosystem. It is often used in pre processing to remove anomalous data from the dataset. In supervised learning, removing the anomalous data from the dataset often results in a statistically significant increase in accuracy. Several anomaly detection techniques have been proposed in literature.

Some of the popular techniques are:

• Density-based techniques (k-nearest neighbor, local outlier factor, isolation forests, and many more variations of this concept).
• Subspace-, correlation-based and tensor-based outlier detection for high-dimensional data.
• One-class support vector machines.
• Replicator neural networks., auto encoders, long short-term memory neural networks
• Bayesian Networks.Hidden Markov models (HMMs).
• Cluster analysis-based outlier detection.
• Deviations from association rules and frequent item sets.
• Fuzzy logic-based outlier detection. [**?**]
• Ensemble techniques, using feature bagging,score normalization and different sources of diversity.

Different methods perform differently a lot on the data set and parameters, and methods it may have little systematic advantages over another when compared across many data sets and parameters. Sample Anomaly Detection Problems. These examples show how anomaly detection might be used to find outliers in the training data or to score new, single-class data. Algorithm for Anomaly Detection. Oracle Data Mining supports One-Class Support Vector Machine (SVM) [**?**]for anomaly detection. When used for anomaly detection, SVM classification does not use a target. Capgemini uses an anomaly detection solution powered by Google Cloud Platform [**?**] which makes it possible to find out threat due to presence of malicious users in advance. With a combination of existing rules-based model and advanced unsupervised machine learning capabilities this is achieved. And provide clients with a more robust and comprehensive solution to track anomalies in run time.

## 3.8 Building a classifier

The success of a classifier meant to be used in practice depends ultimately on its ability to perform well not only with the data used for its testing, but most of all with real-world data. Although a methodologically sound procedure ensures that the classifier is tested against data it has never seen before (i.e., in the training phase), there is no guarantee whatsoever that testing data will be representative of, and resemble closely, real-world data. Such data surely are not available at the time of testing. On the other hand, when using a portion of the training data as a validation dataset to verify and refine the selection of model hyper-parameters, training data could “leak” some information about the validation conditions if the partitioning of data into training and validation sets is not done properly so as to ensure that instances relative to the same (or very close) conditions go in the same part. Connections recorded with the same conditions might, in fact, share some.Because attacks often occur across different tenants, various anomaly detection techniques can combine AI algorithms to analyze attack sequences that are reported on each subscription. These techniques identify the attack sequences as well as prevalent alert patterns, instead of just being incidentally associated with each other.

# Chapter 4

# Second Contribution

We have used KDD cup data set to implement the algorithms for anomaly detection and find vulnerabilities in cloud system. The python packages used to implement this are following

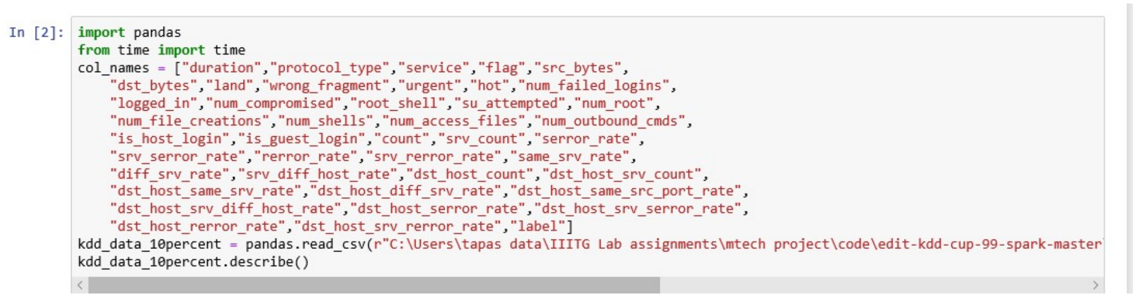
- numpy
- pandas
- scikit-learn
- matplotlib

We started by working on a reduced data set (the 10 percent dataset later). The module numpy allows them to perform high levels of computation to other modules. Pandas module helps to read data files of various formats in order to store them as data frame objects, and it is a popular framework for data science in general. These data frames hold data entries in a fashion a similar to arrays and can be thought of as a table of values. Matplotlib is a Python library that allows us to plot data. Finally, scikit-learn is a package that allows us to apply various machine learning models to data sets as well as it provides tools for data analysis.A tool called Anaconda Navigator was used to implement this.

### 4.0.1 Data Description

**Data Files:**Description of files used from data set
kddcup.name A list of features.
kddcup.data.gz The full data set (743 mb uncompressed)
kddcup.data_10percent.gz A 10% subset of original dataset.Was used to train the classifiers.

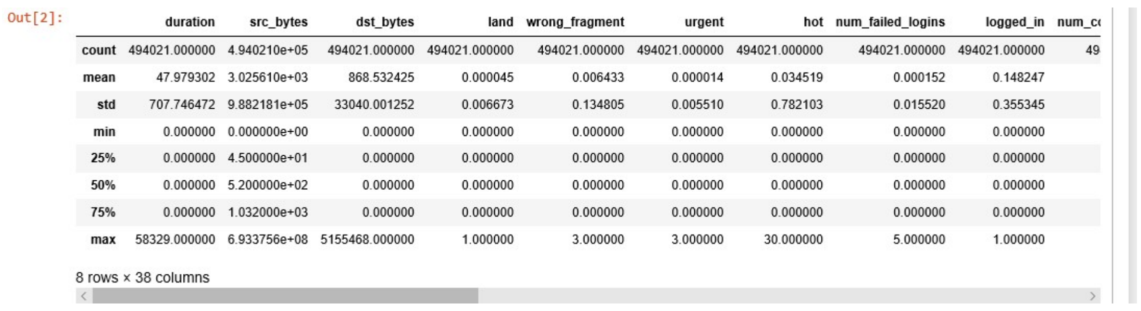
kddcup.testdata.unlabeled_10_percent.gz corrected.gz Test data with corrected labels. training_attack_types A list of intrusion types.



```
In [2]: import pandas
        from time import time
        col_names = ["duration","protocol_type","service","flag","src_bytes",
            "dst_bytes","land","wrong_fragment","urgent","hot","num_failed_logins",
            "logged_in","num_compromised","root_shell","su_attempted","num_root",
            "num_file_creations","num_shells","num_access_files","num_outbound_cmds",
            "is_host_login","is_guest_login","count","srv_count","serror_rate",
            "srv_serror_rate","rerror_rate","srv_rerror_rate","same_srv_rate",
            "diff_srv_rate","srv_diff_host_rate","dst_host_count","dst_host_srv_count",
            "dst_host_same_srv_rate","dst_host_diff_srv_rate","dst_host_same_src_port_rate",
            "dst_host_srv_diff_host_rate","dst_host_serror_rate","dst_host_srv_serror_rate",
            "dst_host_rerror_rate","dst_host_srv_rerror_rate","label"]
        kdd_data_10percent = pandas.read_csv(r"C:\Users\tapas data\IIITG Lab assignments\mtech project\code\edit-kdd-cup-99-spark-master
        kdd_data_10percent.describe()
```

**Figure 4.1:** Initializing the dataset to feed in algorithm

Then the generated output which is generated by reading above variables.



```
Out[2]:
```

| | duration | src_bytes | dst_bytes | land | wrong_fragment | urgent | hot | num_failed_logins | logged_in | num_c |
|---|---|---|---|---|---|---|---|---|---|---|
| count | 494021.000000 | 4.940210e+05 | 494021.000000 | 494021.000000 | 494021.000000 | 494021.000000 | 494021.000000 | 494021.000000 | 494021.000000 | 49 |
| mean | 47.979302 | 3.025610e+03 | 868.532425 | 0.000045 | 0.006433 | 0.000014 | 0.034519 | 0.000152 | 0.148247 | |
| std | 707.746472 | 9.882181e+05 | 33040.001252 | 0.006673 | 0.134805 | 0.005510 | 0.782103 | 0.015520 | 0.355345 | |
| min | 0.000000 | 0.000000e+00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.000000 | 4.500000e+01 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 50% | 0.000000 | 5.200000e+02 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 75% | 0.000000 | 1.032000e+03 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| max | 58329.000000 | 6.933756e+08 | 5155468.000000 | 1.000000 | 3.000000 | 3.000000 | 30.000000 | 5.000000 | 1.000000 | |

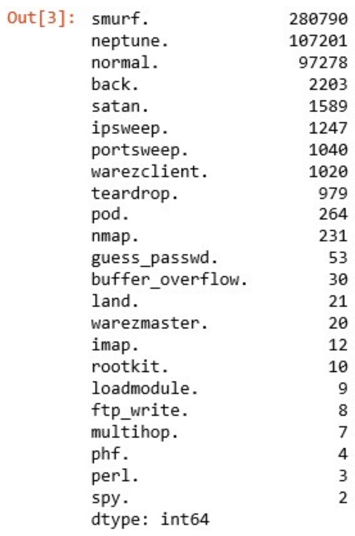
8 rows × 38 columns

**Figure 4.2:** Output table generated after reading data

Now we have our data loaded into a “Pandas“ data frame. In order to get familiar with our data, let’s have a look at how the labels are distributed. kdd_data_10percent[’label’].value_counts() We get following attack types by reading the known attacks from the given dataset.

### 4.0.2 Feature selection

Initially, we will use all features. We need to do something with our categorical variables. For now, we will not include them in the training features. Classification is an important decision-making tools in data mining where the objective is to find some mechanism in order to get a near approximate answer for a particular problem. The answer may be a decision, predicting a behavior, the outcome of a process, or an information, which can act as an input to the next step . By finding an answer which is near to an approximate answer we can conclude that classification method normally gives a near optimal (based on the implementation) solution out of many possible solutions for a particular problem. Once the log files are of an appropriate



```
Out[3]: smurf.              280790
        neptune.            107201
        normal.              97278
        back.                 2203
        satan.                1589
        ipsweep.              1247
        portsweep.            1040
        warezclient.          1020
        teardrop.              979
        pod.                   264
        nmap.                  231
        guess_passwd.           53
        buffer_overflow.        30
        land.                   21
        warezmaster.            20
        imap.                   12
        rootkit.                10
        loadmodule.              9
        ftp_write.               8
        multihop.                7
        phf.                     4
        perl.                    3
        spy.                     2
        dtype: int64
```

**Figure 4.3:** Attack types

format, several statistics are calculated concerning them. Anomaly detection from KDD Data Set performed very well.The log files from different different operating systems create different kind of inconsistencies in the output. One problem is marking false positives in the data set.

Sklearn is a tool that helps dividing up the data into a test and a training set.

```
from sklearn.model_selection import train_test_split

features_train, features_test, labels_train, labels_test = train_test_s
features, labels,
test_size=0.20, random_state=42)
```

Once the data is separated into test and training sets, we can begin to choose a classifier.

```
# import
from sklearn.neighbors import KNeighborsClassifier
```

KNearestneighbor can use any of the following algorithms “auto“, “ball_tree“, “kd_tree“, “brute“,the default is “auto“. We can use any different classifier here other than RandomForestClassifier sklearn is a toolkit which has various algorithms implemented and any of them can be choosen to implement a classifier depending upon what you want to do following shows implementation of RandomForestClassifier.

```
# initialize
clf = RandomForestClassifier()

# train the classifier using the training data
clf.fit(features_train, labels_train)

# compute accuracy using test data
acc_test = clf.score(features_test, labels_test)

print ("Test Accuracy:", acc_test)
```

A classifier can fail to predict the correct result from a given data set if it encounters false positives, and false negatives. A false positive is where something is assumed to be true when it's really false. A false negative is where something is assumed to be false when it's really true. In such scenarios, precision and recall are used to describe the results mathematically rather than false positives and false negatives. Which are defined as Formally, precision and recall are calculated like so:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{4.1}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{4.2}$$

$$Accuracy(F1score) = 2 \times \frac{Precesion \times Recall}{Precision + Recall} \tag{4.3}$$

Precision gives us a measure of how many type of positive classifications are correct where as recall gives us a measure of how correctly we have identified the anomaly.

Sklearn has builtin functions to calculate the precision and recall scores

```
from sklearn.metrics import recall_score, precision_score

precision = precision_score(labels_test, pred, average="weighted")
recall = recall_score(labels_test, pred, average="weighted")

print ("Precision:", precision) #
print ("Recall:", recall) #
```

To detect anomalies using SVM following is the required pseudo code for immplementation,using sklearn's support vector classifier lines of code;

```
from sklearn.svm import SVC
clf = SVC
```

The accuracy of detection and results is different for different algorithms and different approaches. Finally we can conclude that Attacks fall into four main categories:

- DOS: denial-of-service, e.g. syn flood;
- R2L: unauthorized access from a remote machine, e.g. guessing password;
- U2R: unauthorized access to local superuser (root) privileges, e.g., various "buffer overflow" attacks;
- probing: surveillance and other probing, e.g., port scanning.

# Chapter 5

# Conclusion

As we can see anomaly detection is an important topic for research. There is always a scope of improvement in existing methods. This work explored the topic from cloud computing perspective. When you have thousands of virtual machines running inside the data center of cloud service provider then there has to be some mechanism by which they can detect anomalies in their system. Implementation of each service provider would be different because they have different log files and different approaches for implementation. The accuracy of result depends upon the data available and more the system is subject to exposure more it gains intelligence.Existing work is just a small contribution in the field of anomaly detection and shows how anomaly detection techniques can be applied to detect anomalies in a cloud based scenario. Anomaly detection is about finding patterns that do not belong to what is a normal observation.It has its use in each domain of business not only in cloud computing. Using anomaly detection we can detect if a specific port or a service is having abnormal transactions.This kind of information which is available in log files can give vital clue to cyber experts.Anomaly detection is an alternative approach than a traditional intrusion detection system.

# Author’s Biography

Tapas received his B. Tech & MBA dual degree in Information Technology from ABV-IIITM, in 2009. He has been pursuing M. Tech at the Department of Computer Science and Engineering, IIIT Guwahati, since July 2018.