BBB

koeradoera CCC

BBB

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

of

Master

by

koeradoera CCC

under the supervision of

Dr. AAAAA BBB

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY GUWAHATI

Alp 1988

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. Some random thing

b. Some random thing

c. Some random thing

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

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

e. Some random thing

f. Some random thing

koeradoera CCC

ACKNOWLEDGMENTS

Some random thing Some random thing

ABSTRACT

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

1.1 Motivation

Some random thing Some random thing

1.2 Objective of the thesis

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

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

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

2.1 Deployment Models of Clouds

Some random thing Some random thing

Some random thing .

Private Cloud: Some random thing

Public Cloud: Some random thing

Community Cloud: Some random thing

Some random thing .

Hybrid Cloud:Some random thing Some random thing

2.1.1 Cloud Service Categories

Some random thing -

2.1.1.1 IaaS

Some random thing Some random thing

2.1.1.2 PaaS

Some random thing Some random thing

2.1.1.3 SaaS

Some random thing Some random thing

2.1.2 Security Concerns

Some random thing Security Issues with respect to Cloud Computing [12] Some random thing :

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

random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing [13] to launch such attacks. Hence from security perspective it is important to identify such an issue that may exist in underlying cloud infrastructure.

In [14] Some random thing

In [15] Some random thing

In [16] Some random thing [16]. Some random thing In September

2016 [17] Some random thing

2.2 Cloud Services Comparison

Some random thing Some

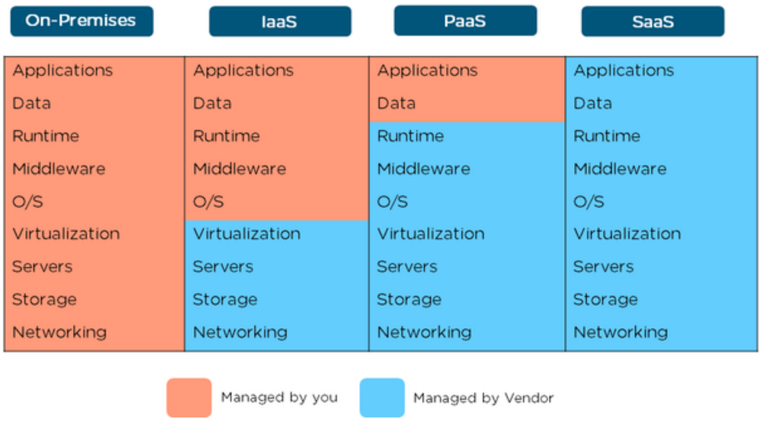

Figure 2.1: Cloud Services Comparison [1]

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

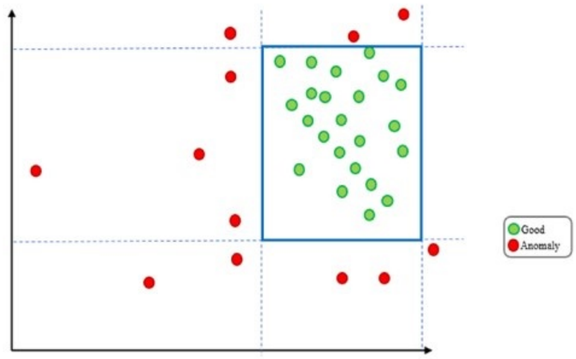

Figure 2.2: Simple Anomaly [2]

Some random thing [18]

2.3.1 Global anomaly

Some random thing Some random thing

Some random thing Some random thing

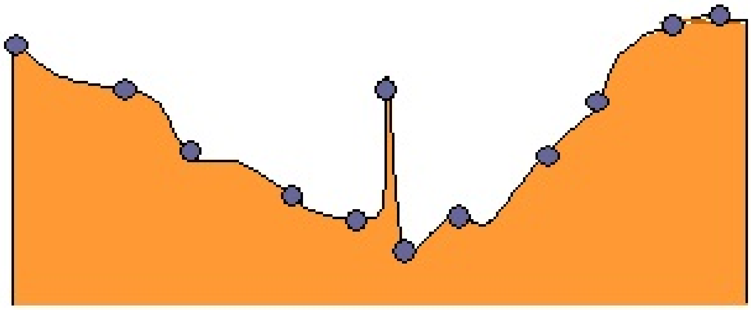

Figure 2.3: Global Anomaly [3]

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 random thing Some random thing 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

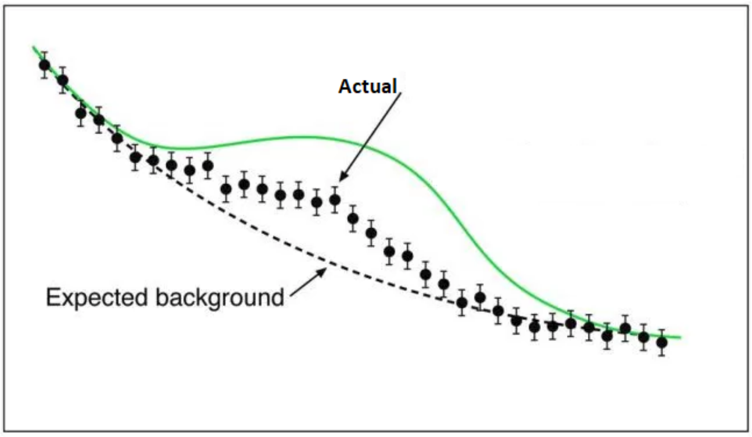

Figure 2.4: Contextual Anomaly Example [4]

Some random thing Some random thing

2.3.3 Collective Anomaly

Some random thing Some random thing

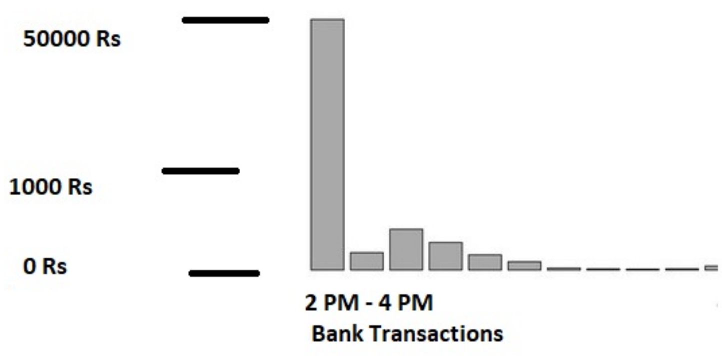

Figure 2.5: Collective Anomaly [5]

Some random thing Some random thing

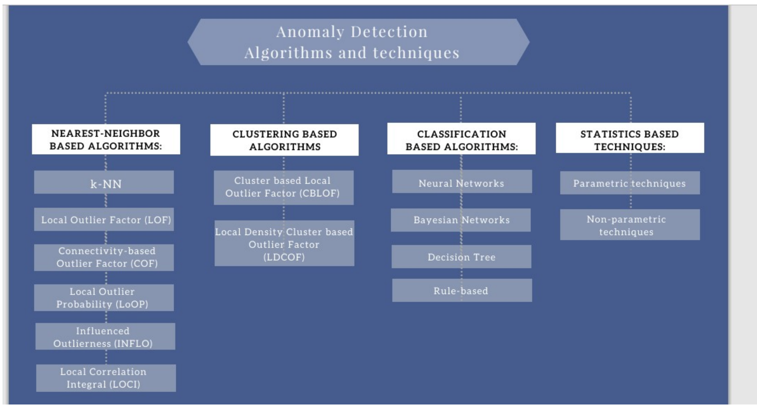

Figure 2.6: Anomaly Detection Algorithms [6]

Some random thing Some random thing

2.3.4 Density-Based Anomaly Detection

Some random thing Some random thing

2.3.5 Clustering-Based Anomaly Detection

Some random thing Some random thing

2.3.6 Support Vector Machine-Based Anomaly Detection

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

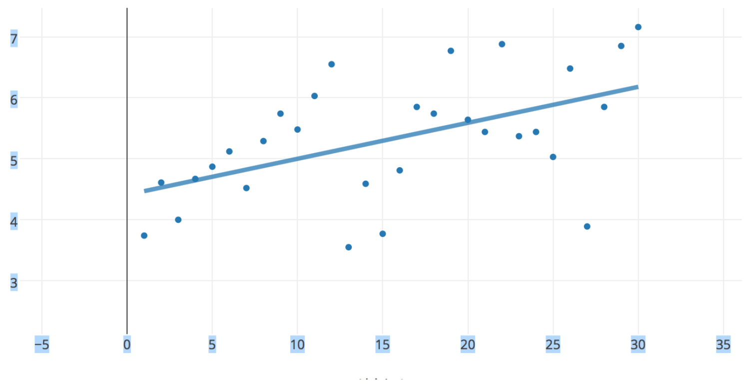

Figure 2.7: Anomaly detection using two variables [7]

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

2.3.7.2 Local Outlier Factor (LOF)

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

2.3.7.4 Support Vector Machine (SVM)

ASome random thing Some random thing

2.3.7.5 Neural Networks Based Anomaly Detection

Some random thing Some random thing

Table 2.1: Anomaly Detection Algorithms comparison [6]

Algorithm	Pros	Cons
• K Nearest Neighbour • K-NN	1. Very easy to understand 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 . Perform poorly on data sets with global anomalies.
K Means	Low Complexity Very easy to implement	Each cluster has pretty equal number of observations Necessity of specifying K Only work with numerical data
Support Vector Machine (SVM)	Find the best separation hyper-plane.Deal with very high dimensional data. Can learn very elaborate concepts. Work very well	Require both positive and negative examples. Require lots of memory. Some numerical stability problems.Need to select a good kernel function
Neural networks based anomaly detection	Learns and does not need to be reprogrammed. Can be implemented in any application	Needs training to operate Requires high processing time for large neural networks The architecture needs to be emulated

Some random thing Some random thing

Chapter 3

asasassas

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

Keywords: Some random thing

3.1 kuttial

Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing Some random thing 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 [19]

3.2 lwoqsas

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

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

Some random thing Some random thing [20] Some random thing Some

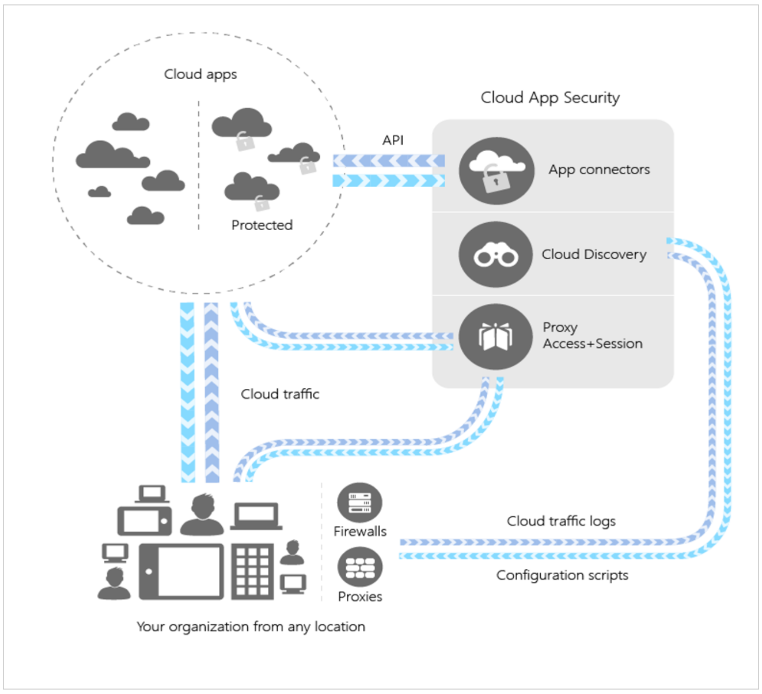

Figure 3.1: Security architecture [8]

random thing Some random thing

3.3 dsdsdsdsds

Some random thing [21] Some random thing
Some random thing Some random thing

1. Intc Threat Intelligence
2. vioral Analytics
3. uscdsdsion Analytics

3.3.1 Idsdssffs

Some random thing Some random thing

3.3.2 middddsds

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

random thing Some random thing .

3.3.3 ddsdssfsfs

Some random thing [22] Some random thing [23] Some random thing

3.4 sfsffssasas

Some random thing [24].Some random thing Some random thing Some random thing Some random thing

Some random thing Events [25], Some random thing

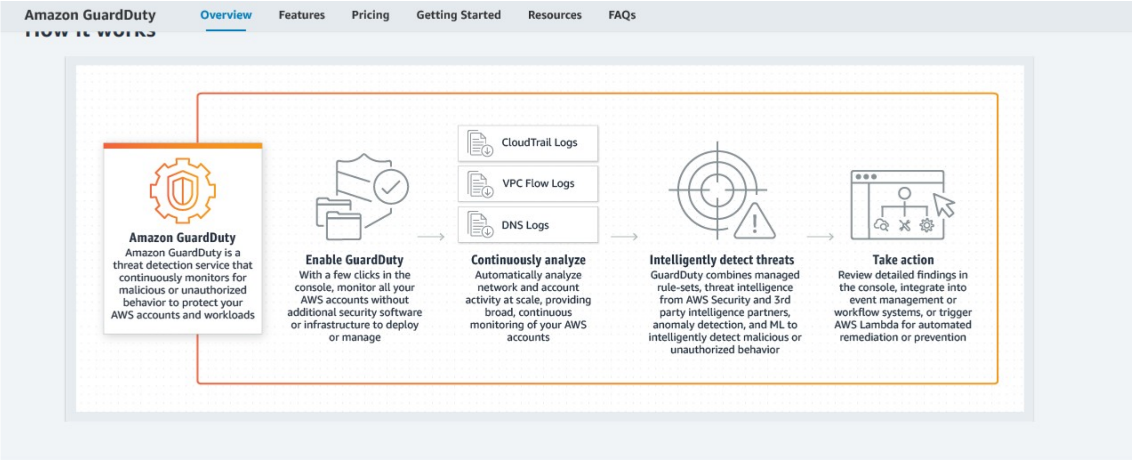

Figure 3.2: Amazon Guard Duty [9]

3.5 dsdssdaaadad

Some random thing [26] 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.

Some random thing Some random thing

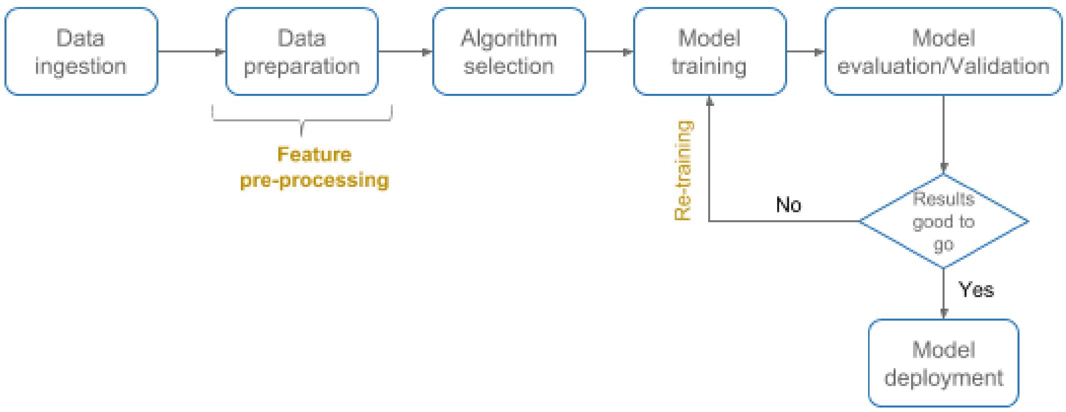

Figure 3.3: GCP Anomaly Detection [10]

3.6 addddadada

Some random thing Some random thing

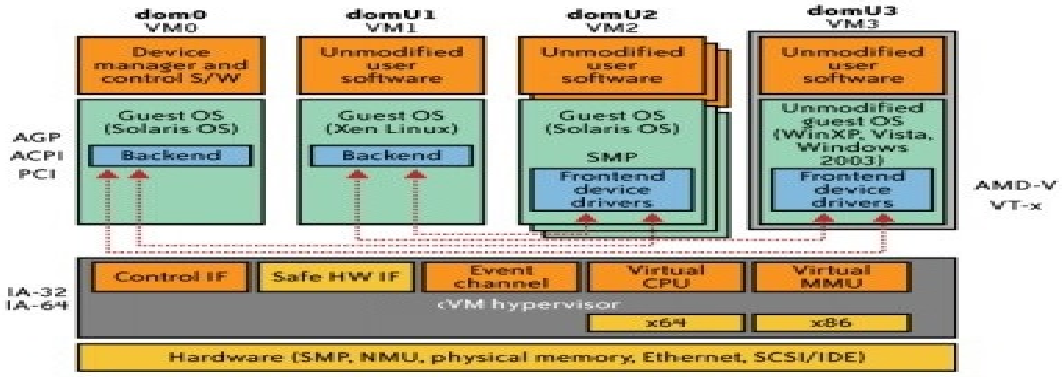

Figure 3.4: Hypervisor working [11]

3.6.1 aadadd

Some random thing (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 Detectinsdd

Some random thing [27], Some random thing [18] Some random thing [28] Some random thing [29]

• 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) Some random thing [30]Some random thing [31] Some random thing

3.8 asaddaddaaa

Some random thing Some random thing

Chapter 4

PPPdsd

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

- numpy
- pandas
- scikit-learn
- matplotlib

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

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

Some random thing Some random thing

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

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

```
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

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

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 Some random thing :

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

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

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

Some random thing Some random thing

```
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) #
```

Some random thing Some random thing

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

Some random thing Some random thing

- 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

Some random thing Some random thing

Bibliography

[1] Namit Kabra. Cloud computing – i (fundamentals). https://namitkabra.wordpress.com/2011/11/14/cloud-computing-i-fundamentals/, November 2011.

[2] Sridhar Alla and Suman Kalyan Adari. *Beginning Anomaly Detection Using Python-Based Deep Learning*. Apress, 2019.

[3] Sayak Paul. Introduction to anomaly detection in python. https://blog.floydhub.com/introduction-to-anomaly-detection-in-python/, apr 2019.

[4] Sergio Santoyo. A brief overview of outlier detection techniques - towards data science. https://towardsdatascience.com/a-brief-overview-of-outlier-detection-techniques-1e0b2c19e561, September 2017.

[5] Daniel Berhane Araya, Katarina Grolinger, Hany El Yamany, Miriam Capretz, and G. Bitsuamlak. Collective contextual anomaly detection framework for smart buildings. 07 2016.

[6] Silvia Valcheva. 5 Anomaly detection algorithms in data mining (with comparison). http://intellspot.com/anomaly-detection-algorithms/.

[7] David Cournapeau. Generalized linear models — scikits.learn v0.6-git documentation. http://scikit-learn.sourceforge.net/0.5/modules/glm.html.

[8] Microsoft. What is cloud app security? microsoft docs. https://docs.microsoft.com/en-us/cloud-app-security/what-is-cloud-app-security.

[9] Amazon. Amazon guardduty – intelligent threat detection - aws. https://aws.amazon.com/guardduty/?nc1=h_ls.

[10] Google. Identify and predict anomalies in firewall rules with forseti. https://cloud.google.com/solutions/partners/forseti-firewall-rules-anomalies.

[11] Ankit Pandey. Virtualization technologies in spi. https://www.ques10.com/p/47790/virtualization-technologies-in-spi-1/.

[12] Diogo AB Fernandes, Liliana FB Soares, João V Gomes, Mário M Freire, and Pedro RM Inácio. Security issues in cloud environments: a survey. *International Journal of Information Security*, 13(2):113–170, 2014.

[13] Michele De Donno, Alberto Giaretta, Nicola Dragoni, Antonio Bucchiarone, and Manuel Mazzara. Cyber-storms come from clouds: Security of cloud computing in the iot era. *Future Internet*, 11(6):127, 2019.

[14] David S Linthicum. Connecting fog and cloud computing. *IEEE Cloud Computing*, 4(2):18–20, 2017.

[15] Jian Shen, Tianqi Zhou, Debiao He, Yuexin Zhang, Xingming Sun, and Yang Xiang. Block design-based key agreement for group data sharing in cloud computing. *IEEE Transactions on Dependable and Secure Computing*, 2017.

[16] Syed Noorulhassan Shirazi, Antonios Gouglidis, Arsham Farshad, and David Hutchison. The extended cloud: Review and analysis of mobile edge computing and fog from a security and resilience perspective. *IEEE Journal on Selected Areas in Communications*, 35(11):2586–2595, 2017.

[17] Constantinos Kolias, Georgios Kambourakis, Angelos Stavrou, and Jeffrey Voas. Ddos in the iot: Mirai and other botnets. *Computer*, 50(7):80–84, 2017.

[18] Arif Sari. A review of anomaly detection systems in cloud networks and survey of cloud security measures in cloud storage applications. *Journal of Information Security*, 06(02):142–154, 2015.

[19] Daniel Chepenko. A density based algorithm for outlier detection. https://towardsdatascience.com/density-based-algorithm-for-outlier-detection-8f278d2f7983.

[20] Microsoft. Discover and manage shadow it in your network. https://docs.microsoft.com/en-us/cloud-app-security/tutorial-shadow-it.

[21] Microsoft. Microsoft way of detecting threats. https://docs.microsoft.com/en-us/azure/security-center/security-center-alerts-overview.

[22] Microsoft. Advanced multistage attack detection in azure sentinel. https://docs.microsoft.com/en-us/azure/sentinel/fusion.

[23] Cloud smart alert correlation in azure security centre. https://docs.microsoft.com/en-us/azure/security-center/security-center-alerts-cloud-smart.

[24] Amazon. Amazon guard duty. https://aws.amazon.com/guardduty/.

[25] Amazon. What is amazon cloud watch events. https://docs.aws.amazon.com/AmazonCloudWatch/latest/events/WhatIsCloudWatchEvents.html.

[26] Google. Forsetti intelligent agents. https://cloud.google.com/solutions/partners/forseti-firewall-rules-anomalies.

[27] S. Roschke, F. Cheng, and C. Meinel. Intrusion detection in the cloud. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, pages 729–734, Dec 2009.

[28] Liu J.G. Pannu, H.S. and S. Fu. Aad: Adaptive anomaly detection system for cloud computing infrastructures.

[29] Y. Dhanalakshmi and I. Ramesh Babu. Intrusion detection using data mining along fuzzy logic and genetic algorithms. *International Journal of Computer Science & Security*, 8:27–32, 2008.

[30] Wun-Hwa Chen, Sheng-Hsun Hsu, and Hwang-Pin Shen. Application of svm and ann for intrusion detection. *Computers & Operations Research*, 32(10):2617 – 2634, 2005. Applications of Neural Networks.

[31] Capgemini. Anomaly detection with machine learning powered by google cloud. https://www.capgemini.com/resources/anomaly-detection-with-machine-learning-powered-by-google-cloud/.

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.