Naif Arab University for Security Science

College of Computer and Information Security

Network Security Department

Anomaly-Based Network Intrusion detection algorithms

By

Wael A. Al-ghamdi

Supervised by:

Dr. Ahmad Elsherif

Thesis submitted in accordance with the regulations for

The Degree of Master of Network Security

Riyadh

1440 - 2019
# List of Contents

Abstract .......................................................................................................................... a
Acknowledgments ........................................................................................................ d
Keywords ....................................................................................................................... e
List of Tables .................................................................................................................. i
List of Figures ................................................................................................................ k

Chapter 1 Introduction .................................................................................................. 1
  1.1 Overview ............................................................................................................... 1
  1.2 Motivation ............................................................................................................ 2
  1.3 Problem Statement .............................................................................................. 2
  1.4 Proposed Approach/Method ............................................................................... 2
  1.5 Objective ............................................................................................................. 3
  1.6 Contribution ....................................................................................................... 3
  1.7 Thesis Organization ......................................................................................... 4

Chapter 2 Background and Related Work ................................................................... 5
  2.1 IDS Evolution ...................................................................................................... 5
  2.2 IDS Classification ............................................................................................... 8
  2.3 Detection Method on IDS ............................................................................... 8
  2.4 Machine Learning Algorithms ....................................................................... 9
  2.5 Feature Selection/Extraction ....................................................................... 10
  2.6 Related Work ..................................................................................................... 12

Chapter 3 Methodology ............................................................................................... 15
  3.1 Dimensionality Reduction ........................................................................... 15
    3.1.1 Singular Value Decomposition ................................................................ 18
    3.1.2 Independent Component Analysis ...................................................... 18
    3.1.3 Principal Component Analysis ............................................................. 19
  3.2 Machine Learning Algorithms ........................................................................... 20
    3.2.1 Decision Tree ....................................................................................... 23
    3.2.2 Naïve Bayes ......................................................................................... 25
4.10.9 Loss Function .................................................................................................................................61
4.10.10 Distribution Function ....................................................................................................................62
4.10.11 Missing Values Handling .............................................................................................................63
4.10.12 RHO + Epochs 10.0 ......................................................................................................................64
4.10.13 RHO with Reproducible (uses 1 thread) + Epochs 10.0 .................................................................65
4.10.14 Final result with best parameters ..................................................................................................66

Chapter 5 Conclusion ...............................................................................................................................68

5.1 Conclusion ..........................................................................................................................................68
5.2 future work .........................................................................................................................................69

References ..............................................................................................................................................70
List of Tables

Table 4.1 List of NSL-KDD dataset files and descriptions ........................................................... 31
Table 4.2 List of attributes of the NSL-KDD dataset. .................................................................... 32
Table 4.3 value type with the attributes of that type ..................................................................... 34
Table 4.4 Singular Value Decomposition result sample. ............................................................... 35
Table 4.5 Attribute that will undergo the dimensionality reduction process. .............................. 35
Table 4.6 Machine Learning Software ......................................................................................... 38
Table 4.7 all attacks detection results ......................................................................................... 40
Table 4.8 all attacks detection results .......................................................................................... 40
Table 4.9 PCA8-Deep Learning results ....................................................................................... 43
Table 4.10 Neptune detection results ......................................................................................... 44
Table 4.11 Guess_passwd detection results .................................................................................. 44
Table 4.12 Smurf detection results. ............................................................................................. 45
Table 4.13 Satan detection results.................................................................................................. 45
Table 4.14 Back detection results. ............................................................................................... 46
Table 4.15 Pod detection results. ................................................................................................. 46
Table 4.16 Nmap detection results .............................................................................................. 47
Table 4.17 IpSweep detection results. .......................................................................................... 47
Table 4.18 PortSweep detection results. ....................................................................................... 48
Table 4.19 TearDrop detection results. ......................................................................................... 48
Table 4.20 Land detection results. ............................................................................................... 49
Table 4.21 Benchmarking detection results. .................................................................................. 51
Table 4.22 Epochs test results ....................................................................................................... 53
Table 4.23 Reproducible feature results. ....................................................................................... 54
Table 4.24 Reproducible with local random seed results. ............................................................ 55
Table 4.25 Epsilon test results. ..................................................................................................... 56
Table 4.26 Train samples per iteration test results ...................................................................... 57
Table 4.27 loss function test results. ........................................................................................... 58
Table 4.28 Distribution function with reproducible test results. ................................................. 59
Table 4.29 Missing Values Handling test results. ......................................................................... 60
Table 4.30 Loss Function test results. .......................................................................................... 61
Table 4.31 Distribution Function test results. ................................................................. 62
Table 4.32 Missing Values handling test results.............................................................. 63
Table 4.33 RHO test results. ......................................................................................... 64
Table 4.34 RHO with Reproducible test results. ............................................................ 65
Table 4.35 Final test results. ......................................................................................... 66
Table 4.36 before and after sensitivity analysis. ............................................................. 67
List of Figures

Figure 2.1 IDS fields of study .................................................................................................................. 7
Figure 3.1 attribute selection and dimensionality reduction ................................................................. 16
Figure 3.2 RapidMiner Dimensionality Reduction process ................................................................. 17
Figure 3.3 Principal Component Analysis representation ...................................................................... 19
Figure 3.4 machine learning train and test model .................................................................................. 21
Figure 3.5 RapidMiner model process .................................................................................................. 22
Figure 3.6 Decision Tree representation .............................................................................................. 23
Figure 3.7 RapidMiner Decision Tree model ......................................................................................... 24
Figure 3.8 Naïve Bayes representation .................................................................................................. 25
Figure 3.9 RapidMiner Naïve Bayes model ........................................................................................... 26
Figure 3.10 Deep Learning representation ............................................................................................ 27
Figure 3.12 RapidMiner Deep Learning model ...................................................................................... 28
Figure 3.13 RapidMiner Comparison process model ............................................................................. 29
Figure 4.1 Presentation of the process of the Sensitivity Analysis ......................................................... 52