

# **Dedication**

This work is dedicated to my family  
To my husband, my daughters and sons  
To my supervisor Dr. Mohamed Awad

## **Acknowledgements**

To Dr. Nouredien Abdelrahman, who introduced me to this field of  
.research, for his aid and encouragement

To my supervisor Dr. Mohammed Awad Elshaikh for his valuable  
.observations, useful suggestions and beneficial criticism

To my uncle engineer Abdelgader Mohammed Ahmed for his deep  
.remarkable continuous scientific help

To my colleague Dr. Eihab Bashier for his practical cooperation and  
.effective support

A special thanks to all the members of my family who shouldered me  
and, for their unconditional love, support, and encouragement through  
.this battle

## Abstract

Intrusion detection is an exemplary method designed to monitor the actions happening in a network. Then analyze them for suspected patterns that may identify a [network](#) or [system](#) violation from someone trying to penetrate and endanger the system. So an Intrusion Detection System (IDS) is software which is applied automatically as a procedure to stop the penetration and attacks of the intruders. It is applied as either Signature recognition or Anomaly detection methodologies. Most of existing IDS required reduction technique in order to minimize the features of data which is irrelevant or redundant. This is needed in case of high dimensionality in network traffic. It is also known that the reduction technique helps the classification algorithms to be very effective. As for the Classification, it achieves and executes the intrusion detection job practically. We realize that Kernel Principal Component Analysis (KPCA) is recognized as a robustification reduction method for standard Principal Component Analysis (PCA) [34]. This research adopts an optimal anomaly detection method to detect multivariate attacks. This method is going to be achieved by measuring the performance of different functions of KPCA as a reduction method applied to different classification algorithms to find out which function of KPCA is the best with any algorithm. Consequently we show that KPCA's methods will not always outperform standard PCA. The final detection's performance, in fact depends on the used classification algorithm. The experiments with NSL-KDD data set demonstrate that the adopted method achieves **98.048%** in detection rate and **98.261%** in precision with **1.484%** false positive rate, consequently outperforms all the other methods. Moreover the results prove that [PCA & K-Nearest Neighbor] outperform [KPCA (Gaussian) & K-Nearest Neighbor] and [KPCA (Quadratic) & K-Nearest Neighbor]. In addition [PCA and Discriminant Analysis] .[outperform [KPCA (Quadratic) & Discriminant Analysis

## المستخلص

ان اكتشاف المتطفلين هي طريقة مثالية لمراقبة الاحداث التي تتم في الشبكة واختبارها بحثاً عن نماذج مشكوك فيها, و قد تعتبر هذه النماذج مهددات للشبكة اوالنظام من شخص ما يحاول اختراقها وبالتالي يهدد النظام. ان نظام اكتشاف المتطفلين برنامج يطبق اتوماتيكياً كإجراء لايقف الاختراقات وهجوم المتطفلين. وهو يقع ضمن احدى المنهجيتين: منهجية التعرف على التوقيعات أو منهجية اكتشاف الشواذ. نجد ان معظم نظم اكتشاف المتطفلين الحالية تتطلب وجود طريقة تقنية لتقليل ميزات البيانات الزائدة عن المطلوب و ليست ذات علاقة بمجموعة البيانات. وهذه التقنية نحتاجها في حالة الابعاد الفائضة في حركة الشبكة. ان اسلوب التقليل الفني يساعد خوارزميات التصنيف لكي تكون شديدة الفعالية. اما فيما يتعلق بالتصنيف فإنه يحدد وظيفة اكتشاف المتطفلين عملياً. نلاحظ ان نظام KPCA يعتبر تقوية لنظام PCA. هذا البحث يبنى طريقة مثالية لاكتشاف الشواذ في حركة الشبكة لتحديد الهجمات متعددة المتغيرات. هذه الطريقة- يتم انجزها بقياس اداء- وتختلف كPCA كطريقة للتقليل مطبقة على خوارزميات تصنيف مختلفة لمعرفة اي وظائف KPCA هي الافضل ومع اي خوارزمية طبقت. وايضا نوضح ان KPCA لا يتفوق دائما على PCA. و اداء الاكتشاف النهائي يعتمد على خوارزمية التصنيف المستخدمة.

كل التجارب اجريت باستخدام برنامج ماتلاب ومجموعة البيانات NSL KDD وهي متوفرة عالمياً. هذه التجارب وضحت أن الطريقة المتبناة تحصلت على 98.048 % في نسبة اكتشاف المتطفلين , 98.261 % في الدقة و 1.484% في نسبة الخطأ الايجابي بذلك تفوق على كل الطرق الأخرى. بالاضافة الى ذلك النتائج تثبت ان نظام

[PCA & KNN] يتفوق على نظام [KPCA (Gaussian) & KPCA (Quadratic) & KNN] و [KNN] وايضاً نظام [PCA and DA] يتفوق على نظام [KPCA (Quadratic) & DA].

## Table of contents

Subject		Page no
	Dedication	I
	Acknowledgements	II
	Abstract	III
	المستخلص	IV
	List of figures	IX
	List of tables	X
<b>Chapter 1 - Introduction</b>		
1.1	Introduction	1
1.2	Problem Statement	2
1.3	Objective	2
1.4	Methodologies	3
1.5	Scope	3
1.6	Research outline	4
<b>Chapter 2 - Background</b>		

2.1	(Definition of Intrusion Detection system (IDS	5
2.2	(Definition of Intrusion Prevention system (IPS	5
2.3	Uses of Intrusion Detection and prevention system (IDPS) technologies	6
2.4	Intrusion Detection and prevention system IDPS Architecture	7
2.5	Intrusion Detection and prevention Systems Categories	8
2.6	Intrusion Detection and prevention System Methodologies	8
2.7	Description of anomaly objects	10
2.8	Categories of anomaly detection based on the nature of data set	10
2.9	Anomaly Detection techniques	11
2.10	Typical Components of IDPS	12
2.11	Evaluation of Anomaly Detection	13
2.12	Definition of Standard measures for evaluating anomaly detection	14
<b>Chapter 3 - Literatures review</b>		
3.1	Introduction	15
2.3	Literatures review of Distance based techniques	15
3.3	Literatures review of Profiling based technique	17
3.4	Literatures review of Model-based technique	18
3.5	Literatures review of Statistical techniques	19
<b>Chapter 4 - Descriptions of the adopted method</b>		
4.1	Introduction	20

4.2	Distance methods	21
4.2.1	Euclidean distance	21
4.2.2	Canberra distance	21
4.2.3	Mahalanobis distance	22
4.3	The reduction stage by PCA and KPCA	22
4.3.1	PCA Ground rules	24
4.3.2	KPCA Ground rules	25
4.4	The classification stage by KNN classification and Discriminant analysis	26
4.4.1	KNN classification	26
4.4.2	Discriminant Analysis	27
4.5	System Architecture	28
4.6	System Algorithm	29
4.6.1	First stage	29
4.6.1.1	Reduction of the feature of the dataset by PCA	29
4.6.1.2	Reduction of the feature of the dataset by KPCA	30
4.6.2	Second stage	30
4.6.2.1	KNN Classification	30
4.6.2.	Discriminant Analysis	31
<b>Chapter 5- The experiments' results and the discussions</b>		
5.1	Tools of experiments	32

5.1.1	Description of KDD Cup 1999 Data	32
5.1.2	Description NSL-KDD data	32
5.2	The Framework of the experiments	33
5.3	Experiments	34
5.4	Performance Measures	34
5.5	Experimental Results and Discussion	36
<b>Chapter 6 - Conclusion and Recommendation</b>		
6.1	Conclusion	53
6.2	Recommendation for future work	54
	References	55-61



## List of figures

No of figure	Title of figure	No of page
Figure 1	Architecture of the proposed system	29
Figure 1.1	Curves of Precision & Detection Rate versus PC [PCA & KNN classification]	35
Figure 1.2	Curve of False Positive Rate versus PC [PCA & KNN classification]	36
Figure 1.3	[ROC curve of [PCA & KNN classification	36
Figure 2.1	Curves of Precision & Detection Rate versus PC [PCA & DA]	37
Figure 2.2	Curve of False Positive Rate versus PC [PCA & DA ]	38
Figure 2.3	[ROC curve of [PCA & DA	38
Figure 3.1	Curves of Precision & Detection Rate versus KPC [[KPCA(Gaussian) & KNN classification	39
Figure 3.2	Curve of False Positive Rate versus KPC [KPCA(Gaussian) & KNN classification]	40
Figure 3.3	ROC curve KPCA (Gaussian) & KNN classification	40
Figure 4.1	Curves of Precision & Detection Rate versus KPC [KPCA (Gaussian) & DA]	41
Figure 4.2	Curve of False Positive Rate versus KPC [KPCA (Gaussian) & DA]	42
Figure 4.3	[ROC curve of [KPCA (Gaussian Kernel) & DA	42

Figure 5.1	Curves of Precision & Detection Rate versus KPC [KPCA (The Laplace) & KNN Classification]	43
Figure 5.2	Curve of False Positive Rate versus KPC [KPCA (The Laplace) & KNN Classification ]	44
Figure 5.3	ROC curve of [ KPCA (The Laplace) & KNN [Classification	44
Figure 6.1	Curves of Precision & Detection Rate versus KPC [KPCA (The Laplace Kernel) & [DA	45
Figure 6.2	Curve of False Positive Rate versus KPC [KPCA (The Laplace Kernel) & DA]	46
Figure 6.3	ROC curve of KPCA (The Laplace Kernel) & DA	46
Figure 7.1	Curves of Precision & Detection Rate Versus KPC [KPCA (The Quadratic) & KNN]	47
Figure 7.2	Curve of False Positive Rate versus KPC [KPCA (The Quadratic) & KNN classification]	48
Figure 7.3	ROC curve of [KPCA (The Quadratic) & KNN [classification	48
Figure 8.1	Curves of Detection Rate & Precision versus KPC [KPCA (The Quadratic) & KNN classification]	49
Figure 8.2	Curve of False Positive Rate versus KPC [KPCA (The Quadratic) & KNN classification]	50
Figure 8.3	[ROC curve of [KPCA(The Quadratic) & DA	50
Figure 9	ROC curve of [KPCA(The Laplace) & KNN [Classification	53

## List of table

<b>No of table</b>	<b>Title of table</b>	<b>No of page</b>
Table 1	[The results of applying [PCA and KNN classification	35
Table 2	The result of applying PCA and DA	37
Table 3	The result of applying KPCA (Gaussian) and KNN classification	39
Table 4	The result of applying KPCA (Gaussian) and DA	41
Table 5	The result of applying KPCA (The Laplace) and KNN Classification	43
Table 6	The result of applying KPCA (The Laplace) and DA	45
Table 7	The result of applying KPCA (The Quadratic) and KNN classification	47
Table 8	The result of applying KPCA (The Quadratic) and DA	49
Table 9	Confusion metrics for evaluations of the adopted method	53