

INTRUSION DETECTION SYSTEM USING HYBRID GSA-K-MEANS

BIBI MASOOMEH ASLAHI SHAHRI

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Computer Science (Information Security)

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

JANUARY 2013

This dissertation is dedicated to my beloved mother Batool Hamed Tabei for her
endless support and encouragement.

ACKNOWLEDGEMENT

Special thanks go to my supervisor, Dr. Anazida Zainal, for her help during the course of this research. Without her time and patience much of this work could not have been accomplished.

Besides, I would like to thank my dear friend Saeed Khorashadi Zadeh to support me during hardship moments of doing my research.

Thank to my dear Amalina Farhan, for helping out to translate the abstract to Bahasa language.

ABSTRACT

Security is an important aspect in our daily life. Intrusion Detection Systems (IDS) are developed to be the defense against security threats. Current signature based IDS like firewalls and antiviruses, which rely on labeled training data, generally cannot detect novel attacks. The purpose of this study is to improve the performance of IDS in terms of detection accuracy and reduce False Alarm Rate (FAR). Clustering is an important task in data mining that is used in IDS applications to detect novel attacks. Clustering refers to grouping together data objects so that objects within a cluster are similar to one another, while objects in different clusters are dissimilar. K-Means is a simple and efficient algorithm that is widely used for data clustering. However, its performance depends on the initial state of centroids and may trap in local optima. The Gravitational Search Algorithm (GSA) is one effective method for searching problem space to find a near optimal solution. In this study, a hybrid approach based on GSA and k-Means (GSA-kMeans), which uses the advantages of both algorithms, is presented. The performance of GSA-kMeans is compared with other well-known algorithms, including k-Means and Gravitational Search Algorithm (GSA). Experimental results on the KDDCup 1999 dataset have demonstrated that the proposed method is more efficient in the detection of intrusive behavior than conventional k-Means and standard GSA which shows 80.62% detection accuracy and 7.45% FAR.

ABSTRAK

Keselamatan adalah satu aspek yang penting dalam kehidupan seharian kita. Sistem Pengesanan Pencerobohan (IDS) dibangunkan untuk menjadi pertahanan terhadap ancaman keselamatan. Tandatangan semasa berasaskan IDS seperti firewall dan antivirus, yang bergantung pada data latihan yang dilabel, umumnya tidak dapat mengesan serangan novel. Tujuan kajian ini adalah untuk meningkatkan prestasi IDS dari segi ketepatan pengesanan dan mengurangkan Kadar Penggera Palsu (FAR), mengikut klasifikasi data. Pengelompokan adalah satu tugas yang penting dalam perlombongan data yang digunakan dalam aplikasi IDS untuk mengesan serangan yang tidak diketahui. Pengelompokan merujuk kepada pengumpulan bersama objek data supaya objek di dalam kelompok adalah sama antara satu sama lain, manakala objek di dalam kelompok yang berbeza adalah berbeza. K-means adalah algoritma yang mudah dan berkesan yang digunakan secara meluas bagi pengelompokan data. Walau bagaimanapun, prestasinya bergantung kepada keadaan awal sentroid dan boleh terperangkap dalam optima tempatan. Algoritma Search Graviti (GSA) adalah salah satu kaedah yang berkesan dalam mencari ruang masalah untuk mencari penyelesaian optimum berhampiran. Dalam kajian ini, pendekatan hibrid berdasarkan GSA dan k-means (GSA-kMeans), yang menggunakan kelebihan kedua-dua algoritma, dibentangkan. Prestasi GSA-kMeans dibandingkan dengan algoritma terkenal lain, termasuk k-Means dan Algoritma Search Graviti (GSA). Keputusan eksperimen set data Piala KDD 1999 telah menunjukkan bahawa kaedah yang dicadangkan adalah lebih cekap dalam pengesanan tingkah laku berbanding daripada k-Means konvensional dan standard GSA yang menunjukkan pengesanan ketepatan 80.62% dan 7.45% FAR.

TABLE OF CONTENT

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENTS	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xi
	LIST OF FIGURES	xiii
	LIST OF ABBREVIATION	xv
	LIST OF APPENDIX	xvi
1	INTRODUCTION	
	1.1 Overview	1
	1.2 Problem Background	2
	1.3 Problem Statement	5
	1.4 Aim of the Project	5
	1.5 Project Objective	5
	1.6 Project Scope	6
	1.7 Significant of the Project	6
	1.8 Project Organization	7
2	LITERATURE REVIEW	
	2.1 Introduction	9
	2.2 Intrusion Detection System	9

2.2.1	Network-Based IDS	11
2.2.2	Host-Based IDS	11
2.3	Types of Networking Attacks	12
2.4	Dataset	14
2.5	Categories of IDS	15
2.5.1	Signature-Based IDSs	15
2.5.2	Anomaly-Based IDSs	16
2.6	Anomaly IDS Classifications	16
2.6.1	Supervised Anomaly Detection	18
2.6.2	Semi-Supervised Anomaly Detection	19
2.6.3	Unsupervised Anomaly Detection	19
2.7	Clustering	20
2.8	Clustering in IDS	22
2.9	Clustering Techniques	24
2.9.1	Self-Organizing Maps (SOM)	24
2.9.2	Fuzzy C-means	27
2.9.3	k-Means	30
	2.9.3.1 k-Means Algorithm	31
2.10	Optimization Techniques	36
2.10.1	A Review of Heuristic Algorithm	36
	2.10.1.1 Gravitational Search Algorithm	38
	2.10.1.2 Particle Swarm Optimization	43
	2.10.1.3 GSA versus PSP	46
2.11	Issues in Clustering for IDS	46
2.12	Summary	48
3	RESEARCH METHODOLOGY	
3.1	Introduction	50
3.2	Research Framework	50
3.3	Phase I	52
	3.3.1 Data Preprocessing	52
	3.3.2 Development of k-Means	53
3.4	Phase II	54
3.5	Phase III	54
3.6	Dataset	55

3.7	Measuring Performance and Evaluation Metrics	57
3.7.1	Euclidean Distance	57
3.7.2	Accuracy in IDS	58
3.8	Summary	60
4	IMPLEMENTATION OF K-MEANS IN IDS	
4.1	Introduction	61
4.2	k-Means Structure	61
4.3	Results and Discussion	63
4.4	Summary	68
5	IMPLEMENTATION OF GRAVITATIONAL SEARCH ALGORITHM AND DESIGN OF THE PROPOSED GSA-KMEANS METHOD	
5.1	Introduction	70
5.2	Development of GSA	70
5.3	GSA-Based Classifier	71
5.3.1	Parameters of GSA Classifier	74
5.3.2	Fitness Function	74
5.4	Results and Discussions	76
5.5	Hybrid k-Means Clustering Algorithm	82
5.6	Results and Discussions	84
5.7	Summary	89
6	RESULT AND DISCUSSION	
6.1	Introduction	90
6.2	Experimental Results on k-Means, GSA and GSA-kMeans	90
6.3	Summary	97
7	CONCLUSION	
7.1	Introduction	98
7.2	Concluding Remarks	98
7.3	Contributions	99

7.4	Future Works and Recommendation	101
7.5	Summary	101
REFERENCES		103
APPENDIX A		109

LIST OF TABLE

TABLE NO.	TITLE	PAGE
2.1	The Comparison on the benefits of network and Host-based IDS	12
2.2	Four classes of attacks	13
2.3	Summary of Popular Datasets intrusion Detection Domain	14
2.4	Features of IDS dataset	15
2.5	Summary of pervious work on clustering in IDS	22
2.6	Description of distance calculation methods	35
3.1	Network data feature labels	56
3.2	Typical IDS evaluation metrics	58
3.3	Description of perfomance measured	59
4.1	Key parameters value used in k-Means	63
4.2	Confusion matrix for standard k-Means on test-1	63
4.3	Confusion matrix for standard k-Means on test-2	64
4.4	Classifications resutl for each classes using test-1 and test-2	67
5.1	Key parameters values used in k-Means	74
5.2	Confusion matrix for standard GSA on test-1	77
5.3	Confusion matrix for standard GSA on test-2	77
5.4	Classification result for each classes using test-1 test-2	81
5.5	Key parameters values in GSA-kMeans	84
5.6	Confusion matrix for GSA-kMeans on test-1	84
5.7	Confusion matrix for GSA-kMeans on test-2	85
5.8	Classification results for each classes using test-1 and test-2	88

6.1	The performance of k-Means, GSA, GSA-kMeans using testing dataset	91
6.2	The classification result for each classes test-1 and test-2 by k-Means, GSA, GSA-kMeans	96

LIST OF FIGURE

FIGURE NO.	TITLE	PAGE
2.1	Organization of a generalized intrusion detection system	10
2.2	Classification of the anomaly detection techniques according to the nature of the processing involved in the “behavioral” model considered	18
2.3	System Architecture of SOM	24
2.4	The Self-Organizing Map	26
2.5	Result of Fuzzy C-means clustering	29
2.6	k-Means Pseudo code	32
2.7	Classification for $k-2$	34
2.8	Newton’s gravity law	39
2.9	General principle of GSA	42
2.10	The encoding of the single particle in the PSO initial population	45
3.1	Research frameworks for designing and development of the proposed technique	51
4.1	Pseudo code for standard k-Means algorithm	62
4.2	Accuracy and FAR for Normal cluster using test-1	65
4.3	Accuracy and FAR for Attack cluster using test-1	65
4.4	Accuracy and FAR for Normal cluster using test-2	66
4.5	Accuracy and FAR for Normal cluster using test-2	67
5.1	Pseud code of standard GSA	71
5.2	The schematic of GSA-based classifier	73
5.3	Accuracy and FAR for Normal cluster using test-1	78
5.4	Accuracy and FAR for Attack cluster using test-1	79
5.5	Accuracy and FAR for Normal cluster using test-2	80

5.6	Accuracy and FAR for Attack cluster using test-2	80
5.7	The framework of GSA-kMeans method	83
5.8	Accuracy and FAR for Normal cluster using test-1	86
5.9	Accuracy and FAR for Attack cluster using test-1	86
5.10	Accuracy and FAR for Normal cluster using test-2	87
5.11	Accuracy and FAR for Attack cluster using test-2	88
6.1	Accuracy and FAR for Normal cluster using k-Means classifier on test-1	92
6.2	Accuracy rate and FAR for Normal cluster using GSA classifier on test-1	93
6.3	Accuracy and FAR for Normal cluster using GSA-kMeans classifier on test-1	93
6.4	Accuracy and FAR for Attack cluster using k-Means classifier on test-1	94
6.5	Accuracy and FAR for Attack cluster using GSA classifier on test-1	95
6.6	Accuracy and FAR for Attack cluster using GSA-kMeans classifier on test-1	95

LIST OF ABBREVIATION

ATM	Automated Teller Machine
FSKSM	Fakulti Sains Komputer dan Sistem Maklumat
PDA	Personal Digital Assistant
IDS	Intrusion Detection System
NIDS	Network Intrusion Detection System
WIPS	Prevention System Wireless Intrusion
FAR	False Alarm Rate
DR	Detection Rate
TPR	True Positive Rate
TNR	True Negative Rate
FNR	False Negative Rate
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative
GSA	Gravitational Search Algorithm
PSO	Particle Swarm Optimization
ACO	Ant Colony Search Optimization
GA	Genetic Algorithm
SA	Simulated Annealing
ML	Machine Learning
LVQ	Learning Vector Quantization
GM	Gaussian Mixtures
K-NN	K-nearest neighbor

LIST OF APPENDIX

NO.	TITLE	PAGE
1	Program Source Code	109

CHAPTER 1

INTRODUCTION

1.1. Overview

In recent years, networks have developed into a compound infrastructure and also in a few years, the network`s bandwidth has grown very considerably so that end users have access into a high-speed network. In addition, large number of devices such as laptops, cellphones and Personal Digital Assistants (PDA) are wirelessly connected to Internet so notably Internet has become a mission–critical infrastructure for companies and every user who has access to the Internet. Because of this, security has become a primary concern in today`s world [1].

The major concern with this situation is that the rate of potential attackers who can violate a given system has extremely climbed. One more reason that emphasizes on the importance of computer security is that today mostly sensitive and critical data is stored on computers which is accessible both locally and remotely, for instance; medical records which consist of personal information are stored on computers and can potentially be accessed remotely by a malicious attacker.

What is more is that many businesses use computer systems to perform their functions consequently, a computer system problem can shut down the whole procedure. For example, an online store would not get any customers if their network connection fails.

Actually, providing privacy is an incredibly important topic. It is vital to make assure that the confidentiality and integrity of sensitive data is protected, which the integrity of essential data is not violated, and that the availability of critical systems is guaranteed [2]. The main purpose of computer security is to accomplish all these goals. Although, all security provisions are met, an attacker mostly find a gap, since the existing security mechanisms fail, so it shows that an additional protection is needed. In order to provide an extra layer of defense, intrusion detection systems (IDSs) have been proposed.

Audit data are scanned by IDSs in order to discovery an indication of malicious behavior. When the indicator of probable security destruction is found, the administrator of system is alerted and presented with a report about the incident, and then the system administrator reacts to the report. Intrusion detection system is an application that can detect unwanted traffic on a network or a device. IDS can be a part of installed software or a physical utilization which observer network traffic in order to detect undesirable activity and events like illegal and malicious traffic that violates security policy, and break up conventional use policies [3].

1.2. Problem Background

Intrusion detection systems are classified in several ways that the most common categorization is by detection method. There are two main types of detection methods: misuse detection and anomaly detection [2]. Misuse detection systems apply regular records that clearly model what is not acceptable. Everything that does not match any of the records is permitted.

Anomaly-based systems use a model of normal activity and anything which is not compliance with the model is considered an attack. The instruction of anomaly detector is that it deems all abnormal events are attacks so all attacks produce anomalous events. Misuse-based systems, on the other hand, can merely identify

attacks which are exist in the regular database, and cannot detect a novel attack unless it has some dissimilarity of one of the attacks already in the rule base [4].

The most significant issue with IDSs is that they produce a huge amount of alerts that are not generated by actual attacks. These false alerts are usually stated as false positives. One solution to prevent the high rate of false positive is to attack classification; this means that normal behavior can be identified from abnormal [5].

Clustering is a method which plays a vital role in distinguishing the attacks from events that its main job is to group the similar data together based on the characteristic they possess. The centroids of each cluster are made by the mean values of the specified number of cluster groups. If the centroids change during the process of clustering, there are several methods used to calculate the new centroids. Clustering algorithms can be used in image analysis, pattern recognition, and bio-informatics and in several other fields. The clustering algorithm comprises two steps; the first step is creating the clusters-calculating centroid and the second step shaping the outliers.

Clustering algorithms are categorized into two types [8]. They are:

- i. Partitional clustering
- ii. Hierarchical clustering

In terms of data mining there are several clustering algorithm typically used in IDS that each one has its own benefits and drawbacks. For instance, applications of fuzzy based methods have an optimistic result in IDS but one significant problem with fuzzy clustering is that it is challenging to attain the membership values.

Another example of clustering algorithm named k-Means is a partitional clustering algorithm that is a center-based clustering method. K-Means is the most popular method due to its simplicity into implementation and efficiency in many

cases [12]. Despite the fact that k-Means algorithm is the most common method in clustering, there are three major parts that require improvements:

- i. The number of k (clusters) must be decided before execution.
- ii. Random choosing of the initial start points makes it impossible to obtain reliable results without much iteration of the entire clustering process
- iii. K-Means algorithms' sensitive to the outliers.

In order to increase the reliability of k-Means, several methods are suggested which will be studied in next chapter. The main focus of this study is on finding a suitable initial centroid for each cluster which has a considerable influence on the clustering result. To improve the performance of k-Means, optimization algorithms will be studied. These algorithms are increasingly analyzed by researchers in many different areas [9, 10]. These algorithms solve diverse optimization problems.

Although many different algorithms are suggested to make k-Means more optimal such as ACO [14], PSO [13, 15], etc. A new optimization algorithm built on the law of gravity, namely Gravitational Search Algorithm (GSA) is offered [11]. This algorithm is based on the Newtonian gravity: “Every particle in the universe attracts every other particle with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them”. All the objects influence each other by the gravity force and it causes a global movement towards the heavier masses. Each mass presents a solution, and the algorithm is navigated by accurately adjusting the gravitational and inertia masses. By the gap of time, masses are attracted by the heaviest mass. This mass will present an optimum solution in the search space [17]. As a result, GSA is a good candidate to be implemented into k-Means algorithm in order to enhance the functionality of finding an initial centroid for each cluster.

1.3. Problem Statement

K-Means clustering algorithm is a popular technique to classify the normal and abnormal behavior in order to increase the detection rate and reduce False Alarm Rates in IDS. Also Gravitational Search Algorithm (GSA) is an optimization algorithm based on interaction of masses in the universe via Newtonian gravity law. These two algorithms both have some weaknesses for example conventional k-Means is highly dependent on the initial state of centroids and it may converge to the local optima rather than global optima so it makes the clustering results generate great uncertainty, and it is important for GSA to create a good initial population because the performance of GSA and most of the population-based algorithms are affected by the quality of the initial population. To improve the performance of clustering by k-Means, a hybrid approach (GSA-kMeans) is proposed to use the advantages of these two algorithms.

1.4. Aim of the Project

The purpose of this study is to improve the detection accuracy in IDS by using hybrid k-Means clustering algorithm with Gravitational Search Algorithm.

1.5. Objective of the Project

The objectives of the study are listed below:

- i. To implement k-Means clustering algorithm to classify IDS data.
- ii. To implement GSA algorithm to cluster arbitrary data and develop a new GSA-based clustering algorithm where k-Means clustering is used to seed the initial swarm.

- iii. To propose a hybrid method (KM-GSA) and evaluate its performance in term of detection accuracy, false positive and false negative rate and compare with conventional k-Means.

1.6. Scope of the Project

The study is limited to the following:

- i. Two algorithms (GSA which is an optimization method and K-Means that is a clustering algorithm) will be used in this study.
- ii. The dataset used in this study is KDD Cup1999 IDS dataset.
- iii. The training and testing data that will be used in this study comprises of 5,092 and 6,890. The composition of these sample data maintains the actual distribution of KDD Cup 1999 data.
- iv. Matlab will be used to code k-Means and GSA.
- v. The classification will be based on two classes which are Normal and Attack.
- vi. Performance will be evaluated based on detection accuracy and false positive rate.

1.7. Significant of the Project

By incorporating the advantage of the k-Means algorithm into GSA, the significant of this study is to increases the performance of GSA in the following two ways:

- i. It decreases the number of iterations and function evaluations performed by GSA to find a near global optimum compared to the original GSA alone.

- i. With the advent of a good candidate solution in the initial population, GSA can search for near global optima in a promising search space and, therefore, find a high quality solution in comparison with the original GSA alone.
- ii. Likewise, k-Means algorithm escapes from local optima. These two goals are performed in order to improve the detection accuracy of IDS and also reduce FAR.

1.8. Organization of the Project

This study consists of seven chapters. The chapters are organized according to different works that involved in this study. The detailed organization of this report is described in following paragraphs. This section presents how this report is organize in different chapters.

Chapter 1 of this report consists of overview of the study, problem background, problem statement, objectives, scope and significance of this study.

Chapter 2 of this report presents a review of the literature related to the area of intrusion detection system. It discusses intrusion detection techniques in details that include clustering techniques and optimization methods, different type of algorithms and clustering problem, functionality of algorithms, and several problems in the previous literature regarding the technique and current solution in IDS.

Chapter 3 is consists of wide description on research methodology, which provides a rich discussion about the flow of this research. This includes how the operational and experimental work has been carried out for the study.

Chapter 4 is the discussion on development of k-Means and evaluation the results of k-Means algorithm in term of accuracy and FAR in IDS.

After that, Chapter 5 discusses designs of GSA-based classifier and proposed method in detail. Results of these two methods are evaluated in this chapter.

Chapter 6 includes the comparison of results on the proposed method will be with k-Means and GSA in term of accuracy and FAR in IDS.

Chapter 7 is the conclusion of overall chapter and future works in the related area of intrusion detection system will be discussed in order to provide a better quality in future study. This includes recommendations for further study.

REFERENCE

- [1] Fredrik V, 2006. Real-Time Intrusion Detection Alert Correlation.
- [2] Revision by Tzeyoung Max Wu, 2009. Intrusion Detection System.
- [3] Sabahi, F., Movaghar, A., 2008. Intrusion Detection: A Survey. IEEE, pp. 23–26.
- [4] Han, L., 2011. Using a Dynamic K-Means Algorithm to Detect Anomaly Activities, in: 2011 Seventh International Conference on Computational Intelligence and Security (CIS). Presented at the 2011 Seventh International Conference on Computational Intelligence and Security (CIS), pp. 1049 – 1052.
- [5] Ali A. Ghorbani, P.K. 2005, 102AD. Research on Intrusion Detection and Response: A Survey 84–101.
- [6] Elshoush, H.T., Osman, I.M., 2011. Alert correlation in collaborative intelligent intrusion detection systems—A survey. Applied Soft Computing 11, 4349–4365.
- [7] Denatiou, D.K., John, A., 2012. Survey on data mining techniques to enhance intrusion detection. IEEE, pp. 1–5.
- [8] Jose F. Nieves, 2009. Data Clustering for Anomaly Detection in Network Intrusion Detection.
- [9] Chen, B., Tai, P.C., Harrison, R., Pan, Y., 2005. Novel hybrid hierarchical-K-Means clustering method (H-K-Means) for microarray analysis, in: IEEE Computational Systems Bioinformatics Conference, 2005. Workshops and Poster Abstracts. Presented at the IEEE Computational Systems Bioinformatics Conference, 2005. Workshops and Poster Abstracts, pp. 105 – 108.
- [10] García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., Vázquez, E., 2009. Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security 28, 18–28.
- [11] Hatamlou, A., Abdullah, S., Nezamabadi-pour, H., 2012. A combined

approach for clustering based on K-Means and gravitational search algorithms. *Swarm and Evolutionary Computation*.

- [12] Tian, L., Jianwen, W., 2009. Research on Network Intrusion Detection System Based on Improved K-Means Clustering Algorithm, in: *International Forum on Computer Science-Technology and Applications*, 2009. IFCSTA '09. Presented at the International Forum on Computer Science-Technology and Applications, 2009. IFCSTA '09, pp. 76–79.
- [13] Yang, X., Yuan, J., Yuan, J., Mao, H., 2007. A modified particle swarm optimizer with dynamic adaptation. *Applied Mathematics and Computation* 189, 1205–1213.
- [14] Z. Baojiang, L. Shiyong, 2007. Ant colony optimization algorithm and its application to neuro-fuzzy controller design, *Journal of Systems Engineering and Electronics* 603–610.
- [15] F.V.D. Bergh, A.P. Engelbrecht, 2006 .A study of particle swarm optimization particle trajectories, *Information Sciences* 937–971.
- [16] Rashedi, E., Nezamabadi-pour, H., Saryazdi, S., 2009. GSA: A Gravitational Search Algorithm. *Information Sciences* 179, 2232–2248.
- [17] Wu, S.X., Banzhaf, W., 2010. The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing* 10, 1–35.
- [18] Harley Kozushko, 2003. *Intrusion Detection: Host-Based and Network-Based Intrusion Detection Systems*.
- [19] Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A.A., 2009. A detailed analysis of the KDD CUP 99 data set. *IEEE*, pp. 1–6.
- [20] Smaha, S.E. 1988. Haystack: An intrusion detection system. *Proceeding of the Fourth Aerospace Computer Security Application Conference* (pp.37-44).
- [21] Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection. *ACM Computing Surveys* 41, 1–58.
- [22] Song, X., Wu, M., Jermaine, C., Ranka, S., 2007. Conditional Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering* 19, 631–645.
- [23] Lu, W., Traore, I., 2008. Unsupervised anomaly detection using an evolutionary extension of k-Means algorithm. *International Journal of*

Information and Computer Security 2, 107.

- [24] Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. *ACM Comput. Surv.* 31, 264–323.
- [25] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of Second International Conference on Knowledge Discovery and Data Mining*, E. Simoudis, J. Han, and U. Fayyad, Eds. AAAI Press, Portland, Oregon, 226-231.
- [26] Guha, S., Rastogi, R., and Shim, K. 2000. ROCK: A robust clustering algorithm for categorical attributes. *Information Systems* 25, 5, 345-366.
- [27] ErtÄoz, L., Steinbach, M., and Kumar, V. 2003. Finding topics in collections of documents: A shared nearest neighbor approach. In *Clustering and Information Retrieval*. 83-104.
- [28] Yu, D., Sheikholeslami, G., and Zhang, A. 2002. Findout: Finnding outliers in very large datasets. *Knowledge And Information Systems* 4, 4, 387-412.
- [29] Smith, R., Bivens, A., Embrechts, M., Palagiri, C., and Szymanski, B. 2002. Clustering approaches for anomaly based intrusion detection. In *Proceedings of Intelligent Engineering Systems through Artificial Neural Networks*. ASME Press, 579-584.
- [30] Pachghare, V.K., Kulkarni, P., Nikam, D.M., 2009. Intrusion Detection System using Self Organizing Maps. *IEEE*, pp. 1–5.
- [31] Kohonen, T, 1997. *Self-Organizing Maps*, Springer Series in Information Sciences. Berlin, Heidelberg: Springer.
- [32] P. Lichodziejewski, A. Zincir-Heywood, and M. Heywood, 2002. Dynamic intrusion detection using self organizing maps.
- [33] McHugh, J, (2000). Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by lincoln laboratory. *ACM Trans. on Information and System Security* 262–294.
- [34] Weina Wang, Yunjie Zhang, Yi Li, Xiaona Zhang, 2006. The Global Fuzzy C-Means Clustering Algorithm. *IEEE*, pp. 3604–3607.
- [35] J. An, G. Yue, F. Yu, and R. Li, Springer Berlin / Heidelberg, 2006. Intrusion detection based on fuzzy neural networks. In J. Wang, ZhangYi, J.

- M. Zurada, B.-L. Lu, and H. Yin, editors, *Advances in Neural Networks - Third International Symposium on Neural Networks (ISNN '06)*, volume 3973 of *Lecture Notes in Computer Science*, pages 231–239.
- [36] JAIN, A. K., TOPCHY, A., LAW, M. H. C., & BUHMANN, J. M. 2004. Landscape of clustering algorithms. Pages 260–263 of: *Proceedings of the International Conference on Pattern Recognition*, vol. 1.
- [37] STEINHAUS, H. 1956. Sur la division des corp materiels en parties. *Bulletin of Acad. Polon. Sci.*, IV(C1. III), 801–804.
- [38] LLOYD, S. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28, 129–137.
- [39] BALL, G., & HALL, D. 1965. ISODATA, a novel method of data anlysis and pattern classification. Tech. rept. NTIS AD 699616. Stanford Research Institute, Stanford, CA.
- [40] MACQUEEN, J. 1967. Some methods for classification and analysis of multivariate observations. Pages 281–297 of: *Fifth Berkeley Symposium on Mathematics, Statistics and Probability*. University of California Press.
- [41] S.Z. Selim, K. Alsultan, 1991. A simulated annealing algorithm for the clustering problem, *Pattern Recognition* 1003–1008.
- [42] K.S. Al-Sultan, 1995. A Tabu search approach to the clustering problem, *Pattern Recognition* 1443–1451.
- [43] U. Maulik, S. Bandyopadhyay, 2000. Genetic algorithm-based clustering technique, *Pattern Recognition* 1455–1465.
- [44] K. Krishna, M. Narasimha Murty, 1999. Genetic K -means algorithm, *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)* 433–439.
- [45] A.K. Qin, P.N. Suganthan, 2004. Kernel neural gas algorithms with application to cluster analysis, in: *Proceedings—International Conference on Pattern Recognition*.
- [46] A.K. Qin, P.N. Suganthan, 2004. A robust neural gas algorithm for clustering analysis, in: *Proceedings of International Conference on Intelligent Sensing and Information Processing, ICISIP 2004*.
- [47] C. Ching-Yi, Y. Fun, 2004. Particle swarm optimization algorithm and its application to clustering analysis. in *Networking, Sensing and Control*,

2004 IEEE International Conference.

- [48] P. Jin, Y.L. Zhu, K.Y. Hu, 2007. A clustering algorithm for data mining based on swarm intelligence, in: Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, ICMLC 2007.
- [49] Manas Ranjan Patra, M.P., 2005. Some Clustering Algorithms to enhance the performance of the network intrusion detection system.
- [50] Jain, A.K., 2010. Data clustering: 50 years beyond K-Means. *Pattern Recognition Letters* 31, 651–666.
- [51] O. Cordon, S. Damas, J. Santamarı, 2006. A fast and accurate approach for 3D image registration using the scatter search evolutionary algorithm, *Pattern Recognition Letters*, 1191–1200.
- [52] H. Nezamabadi-pour et al, 2006. Edge detection using ant algorithms, *Soft Computing* 623–628.
- [53] A. Kalinlia, N. Karabogab, 2005. Artificial immune algorithm for IIR filter design, *Engineering Applications of Artificial Intelligence* 919–929.
- [54] K.S. Tang, K.F. Man, S. Kwong, Q. 1996. Genetic algorithms and their applications, *IEEE Signal Processing Magazine* 22–37.
- [55] S. Kirkpatrick et al, 1983. Optimization by simulated annealing, *Science* p.671–680.
- [56] V. Gazi, K.M. Passino, 2004. Stability analysis of social foraging swarms, *IEEE Transactions on Systems, Man, and Cybernetics – Part B* 539–557.
- [57] D. Holliday, R. Resnick, J. Walker, *Fundamentals of physics*, John Wiley and Sons.
- [58] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann. Peddabachigari, S., Abraham , A., Grosan, C., & Thomas, J. (2007). Modeling intrusion detection system using hybrid intelligent systems. *Journal of Network and Computer Applications*, 30,114-132.
- [59] Chebrolu Srilatha, Abraham Ajith, Thomas Johnson. Hybrid feature selection for modeling intrusion detection systems. In: Pal NR, et al , editor. 11th International conference on neural information processing, ICONIP' 04. *Lecture Notes in Computer Science*. Vol. 3316. Germany: Springer Verlag; 2004. P. 1020e5. ISBN 3-540-23931-6.
- [60] Mukkamala et al., 2004b Srinivas Mukkamala, Andrew H. Sung, Ajith

Abraham, Vitorino Ramos Intrusion detection systems using adaptive regression splines, in: I.Seruca, J. Filipe, S. Hammoudi, J. Cordeiro (Eds.), Sixth international conference on enterprise information systems, ICEIS'04, Portugal, vol.3(2004),pp.26-33 ISBN 972-8865-00-7.

- [61] W. Lee, S.J. Stolfo, K.W. Mok, Adaptive intrusion detection: a data mining approach, *Artif. Intell. Rev.* 14(6)(2000)533-567.
- [62] Shelly Xianon Wu, Wolfgang Banzhaf, The use of computational intelligence in intrusion detection systems: A review Memorial University of Newfoundland, St John's, NL A1B3X5, Canada.
- [63] Bahrololoum, A., Nezamabadi-pour, H., Bahrololoum, H., Saeed, M., 2012. A prototype classifier based on gravitational search algorithm. *Applied Soft Computing* 12, 819–825.
- [64] Bahrololoum, A., Nezamabadi-pour, H., Bahrololoum, H., Saeed, M., 2012. A prototype classifier based on gravitational search algorithm. *Applied Soft Computing* 12, 819–825.