Hybrid Intrusion Detection System based on DNA Encoding, Teiresias Algorithm and Clustering Method

Omar Fitian Rashid
Department of Computer Technology Engineering, Al Hikma University College, Baghdad, Iraq.

Mazin S. Al-Hakeem*
Department of Computer Technology Engineering, Al Hikma University College, Baghdad, Iraq.
E-mail: alhakeem.ms@gmail.com

Received August 11, 2021; Accepted November 27, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V19I1/WEB19036

Abstract

Until recently, researchers have utilized and applied various techniques for intrusion detection system (IDS), including DNA encoding and clustering that are widely used for this purpose. In addition to the other two major techniques for detection are anomaly and misuse detection, where anomaly detection is done based on user behavior, while misuse detection is done based on known attacks signatures. However, both techniques have some drawbacks, such as a high false alarm rate. Therefore, hybrid IDS takes advantage of combining the strength of both techniques to overcome their limitations. In this paper, a hybrid IDS is proposed based on the DNA encoding and clustering method. The proposed DNA encoding is done based on the UNSW-NB15 database by dividing the record's attributes into four groups, including State, Protocol, Service, and the rest of the features is Digits. Four DNA characters were used to represent each protocol attribute values. While two DNA characters are used to represent State, Service and Digits attributes values. Then, the clustering method is applied to classify the records into two clusters, either attack or normal. The current experiment results showed that the proposed system has achieved a good detection rate and accuracy results equal to 81.22% and 82.05% respectively. Also, the system achieved fast encoding and clustering time that equal 0.385 seconds and 0.00325 seconds respectively for each record.

Keywords

Intrusion Detection System, DNA Encoding, Clustering Algorithm, UNSW-NB15 Database.

Introduction

The intrusion detection systems (IDS) are used to identify unauthorized users and the misuse of computer systems and networks, where these systems have become a primary

component in computer or network security. Prevention of intrusions entirely depends on the detection capability of these systems. Several IDS methods were introduced and applied including DNA encoding and clustering algorithms. DNA encoding was employed first to convert the used dataset records' characters to a DNA sequence. Various approaches of DNA encoding have been developed for IDS, where these systems achieved high detection rates and accuracy. Al-Ibaisi et al. (Al-Ibaisi et al, 2008) recommended using DNA encoding for IDS; hence their encoding method is done by dividing the record's attributes as static and dynamic. Then, three DNA characters are applied to each value and put three characters as a header in front of static parameters attributes values. Hameed and Rashid (Hameed et al, 2014) introduced a misuse IDS based on DNA sequence. Three steps are followed to perform this system. In the first step, the network traffic is converted to a DNA sequence. In the second step, the attack signature keys and their positions are extracted using the Teiresias algorithm. Teiresias is an algorithm that can detect and report all existing patterns in a set of input sequences without using alignment. In the last step, network traffic has classified either attack or normal based on the attack signature key and their positions using the Horspool algorithm. Clustering is the second method that gathers the objects in distinct groups, yet these groups are called clusters. Roshan et al. (Roshan et al, 2018) proposed a design for IDS, and this system was based on Extreme Learning Machines and clustering, whereas this system can identify both known and other attacks. Horng et al. (Horng et al, 2011) suggested a novel IDS based on the combination of SVM and hierarchical clustering algorithm. They suggested a feature selection method by using the KDDCup dataset. A clustering method is expected to solve the intrusion detection system problem by calculating the distance between two samples only once (Wei et al, 2017). Wang (Wang et al, 2011) improved the K-means clustering for the intrusion detection method that trains on unlabeled data to observe recent attacks.

There are two key techniques for detection: anomaly detection that describes manners that do not match the normal ones. The second one is misuse detection that acts as a matching procedure based on known attack signatures. The most significant drawback of anomaly detection is the high False Alarm Rates, while difficulties in notifying the recent attacks are considered as the misuse detection drawbacks (Mukherjee et al, 2012). Therefore, hybrid IDS that take advantage of combining the strength of both techniques is proposed to avoid the aforementioned weaknesses to enhance the IDS performance. Many studies have been conducted to improve hybrid IDS. In this regard, Modi & Patel (Modi et al, 2013) came up with a novel hybrid IDS for cloud computing using several classifiers, and these classifiers are Associative, Bayesian, and Decision Tree. A new feature selection

method for hybrid IDS is also suggested by Chung and Wahid (Chung et al, 2012) based on rough set theory and particle swarm optimization. Thus, the performance of the hybrid IDS was based on the matching and genetic algorithm (Desai et al, 2016). The matching algorithm applies to distinguish internal attacks, and a genetic algorithm adapted to observe external attacks. Chahal and Kaur (Chahal et al, 2016) proposed the system of hybrid intrusion detection wherein the pattern matching algorithm was implemented for misuse detection while the clustering algorithm applied for anomaly detection. Rustam and Maharanian (Rustam et al, 2020) designed an IDS model based on combining the C-Means method and Laplacian Score methods, where the first method utilized as a classifier the second method used for feature selection. It is reported that the hybrid IDS model produced based on the new clustering method, which is used the classifying mapping, leads to increase model accuracy for new attack types (Bharti et al, 2010). Another hybrid IDS method is introduced by Choudhary and Kesswani (Choudhary et al, 2019) using a routing protocol system to detect routing attacks. Another approach is recommended where an IDS is based on K-means clustering, and Raspberry Pi was applied, and the role of Raspberry Pi is to identify blocks and keep an intruder's address (Sumanth et al, 2020). An industrial IDS is presented by Liang et al. (Liang et al, 2019), and it is application based on data clustering, whereabouts measuring weighted data distances using the priority threshold. A multilayer IDS is suggested by utilizing semi-supervised clustering to solve the unsupervised clustering problem. The suggested system used both genetic algorithm and tag extension method (Wang et al, 2019). Furthermore, an IDS based on automatic clustering was proposed by Shojafar et al. (Shojafar et al, 2019), where this method can detect the similarity between the cluster element and the rest of the cluster. The IDS accuracy and detection rate are enhanced by building a new IDS that depends on the SVM and clustering algorithm (Liang et al, 2019). The clustering algorithms are used to split the data, then applied machine learning to model these parts. Chen (Chen et al, 2019) built a multilayer IDS model, where this model was built based on the clustering method and neural network. Rustam and Talita (Rustam et al, 2018) proposed IDS based on fuzzy kernel robust clustering. This system was applied to the KDDCup data set and classify the data into five clusters (one cluster for normal records and the other four clusters are for different attack types). Additionally, A training model for IDS based on machine learning classification was built (Verma et al, 2018). This system can be applied with or without the clustering method. Eslamnezhad and Varjani (Eslamnezhad et al, 2014) introduced IDS based on the K-means clustering method, this method overcomes the sensitivity of initial centers in the K-means algorithm and increases clustering quality. An IDS is based on the clustering method using a genetic algorithm, where each k cluster centroid is with a chromosome (Aissa et al, 2015). A

multi-agent architecture is considered the best tool concerning IDS based on ant colony clustering, and it can find the new attacks (Abdurrazaq et al, 2014).

In the current paper, the review of the various developments in this field encourages the author to build a hybrid IDS utilizing the DNA encoding method for clustering and classify data into different classes. This method is applied using a UNSW-NB15 database, and the outcome is adding new security tools.

Research Method

The suggested approach is done via three phases. These phases are building a DNA encoding method, extracting keys and their positions (where this position is equal to the first character location for extracted key), and then applying the clustering method. In the DNA encoding phase, the plaintext is converted to the DNA sequence. It is used as an attempt to develop a DNA sequence for the used data set records and management to achieve better detection results. This is a well-known procedure utilized in different fields of the computer system such as cryptography, steganography, digital signature, and others.

In the present work, the UNSW-NB15 data set is used for training and checking. This data set consists of nine different attack types and regular activities, and each record of this data set contains 49 features (Moustafa et al, 2015). These features are placing into four distinct groups, namely, State, Protocol, Service, and Digits. Table 1 explains the UNSW-NB15 dataset features and their values.

Table 1 UNSW-NB15 dataset features categorization

Name	Values
State	7 values
Protocol	131 values
Service	13 values
The rest features (Digits)	11 values

The DNA encoding functions is to transform all attribute values into DNA sequences. These attributes are arranged into four groups, as illustrated in Table 1. The state attributes have seven different symbolic values and two DNA characters that can treat and describe all state attribute values. The protocol attribute has 131 different symbolic values and four DNA characters and deal with and display all protocol attribute values. The service attribute has 13 different nominal values, and two DNA characters can use and describe all service attribute values. The rest of the attributes (digits) have 11 different numerical values, and two DNA characters are run to perform all these values. The DNA

sequences for all attribute values formed by the present work are expressed in Tables 2, 3, 4, and 5. The same DNA sequences can be used for different attribute values have no consequences on the results s because it is not used for encryption, and these DNA sequences are used for extraction keys and positions only.

Table 2 DNA sequences for state attribute values

State	DNA Sequence	State	DNA Sequence
ACC	GA	INT	CT
CLO	AT	REQ	TC
CON	CA	RST	AG
FIN	CC		

Table 3 DNA sequences for protocol attribute values

Protocol	DNA Sequence	Protocol	DNA Sequence	Protocol	DNA Sequence
3pc	CCGC	merit-inp	AGTA	snp	AGTT
a/n	GAAG	mfe-nsp	TATG	sprite-rpc	GCCG
aes-sp3-d	ACTC	mhrp	GTGA	sps	GCCA
Any	TCCT	micp	TACT	srp	TAAT
Argus	GGCA	mobile	ACCG	st2	TAAA
Aris	GGTT	mtp	CGAA	stp	TCCG
Arp	AGTC	mux	ATTC	sun-nd	TGTA
ax.25	GGGC	narp	CTTC	Swipe	CGCA
bbn-rcc	CCCA	netblt	ACCA	iso-ip	CAAA
Bna	ATAC	nsfnet-igp	GATC	iso-tp4	GCAC
br-sat-mon	CCAC	nvp	ACCT	kryptolan	GTCA
Cbt	CGGA	ospf	TTAA	l2tp	ACAT
Cftp	TGGC	pgm	TTGG	larp	GCCT
Chaos	ATCT	pim	TTCC	leaf-1	CAGT
compaq-peer	GTAA	pipe	ACGC	leaf-2	CGCT
Cphb	AACA	pnni	GCGA	Tcf	GGAA
Cpnx	AGGG	pri-enc	GATT	Tcp	CCGG
Crtp	TAAG	prm	TTCT	Udp	CTGT
Crudp	CCAG	ptp	TCAG	Unas	TGGA
Dcn	TGAT	pup	ACGA	Uti	TCGG
Ddp	CAGA	pvp	CCTC	Wsn	ATGG
Ddx	TACC	qnx	GGCT	Xnet	GAGC
Dgp	TTCG	rdp	GGAG	xns-idp	AACG
Egp	CAAC	rsvp	GCTC	Xtp	AAAA
Eigrp	TCAC	rvd	GTCC	Zero	GGGG
Emcon	ATGT	sat-expak	ACTT	idpr	CGTC
Encap	TCGA	sat-mon	GGCG	idpr-cmtp	AGCT
Etherip	GCTA	sccopmce	TTAG	idrp	CAGG
Fc	TGCA	scps	CACG	ifmp	GTAT
Fire	CATA	sctp	CCTT	igmp	ATAA
Ggp	ACGT	sdrp	GGGA	igp	TCCA
Gmtp	GCTT	secure-vmtp	ACAA	il	AAAG
Gre	GGTC	sep	CGGT	i-nlsp	GGAC
Hmp	CCGA	skip	ACAC	ipv6-no	GTGC
Iatp	CGTG	sm	GAAA	ipv6-opts	GATG
Ib	CGAG	smp	TATC	ipv6-route	TAGT
trunk-1	TCGC	Vmtp	CCCT	ipx-n-ip	TCTT
trunk-2	TCTA	Vrrp	ACAG	irtp	GTTC
Ttp	CACC	wb-expak	GAAT	isis	ATTG
Vines	TGGT	wb-mon	TCCC	Tlsp	CGCG
Visa	CGGG	tp++	GACA		

Table 4 DNA sequences for service attribute values

Service	DNA Sequence	Service	DNA Sequence
-	GG	pop3	TA
Dhcp	GC	radius	TC
Dns	CT	smtp	GT
ftp	AT	snmp	AG
ftp-data	TG	Ssh	CC
http	AC	Ssl	GA
Irc	TT		

Table 5 DNA sequences for digits attributes values

Digit	DNA Sequence	Digit	DNA Sequence
0	TA	6	GC
1	CT	7	GT
2	CG	8	CA
3	AG	9	GA
4	TT	.	CC
5	TG		

An example of converting one record to DNA sequences based on DNA encoding method is exhibited as follow. The record is:

47933,0.000009,unas,-,INT,2,0,200,0,111111.1072,254,0,88888888,0,0,0,0.009,0,0, 0,0,0,0,0,0,0,0,100,0,0,0,8,2,5,4,4,8,0,0,0,13,8,0",

and its equivalent DNA sequences is:

TTGTGAAGAGTA
CCTATATATATAGATGGAGGCTCGTACGTATATACTCTCCTCTCTCCCTTGTCG
CGTGTTTACAC
ACACACACACACATATATATACCTATAGATATATATATATATATATATACTTA
TATATATACACGTGTTTTCATATATACTA GCATA.

The second phase of the proposed approach involves the Teiresias algorithm; this algorithm is utilized to identify all existent patterns from the DNA sequences set. The Teiresias algorithm refers to a training database to obtain keys and their positions for the hybrid method. This is performed by extracting keys and their positions for misuse method that depends on attacks signatures (blocks of attacks records that diverge from blocks of typical records). Then, extract keys and their positions for an anomaly method depend on user behaviors (blocks of normal records that vary from blocks of attack records). The extracted keys and their positions for both misuse and anomaly method are proved as in Table 6 and Table 7, subsequently.

Table 6 Misuse method keys and its positions

Keys	Positions Numbers
AGCCG	134, 58, 240 and 222
GCTTG	60, 122, 26 and 108
CTCCC	53
TCCCT	54 and 78
ACGCT	102 and 12

Table 7 Anomaly method keys and its positions

Keys	Position Number
AGCTG	90
TTGTA	120
TACTA	149
GTTAT	109
TATTT	117
GATAT	27

In the third phase of the proposed approach, 4000 random testing records converted to DNA sequences are used. This is followed by classifying these records either to a normal cluster or attack cluster, based on the two extracted sets of keys and their positions as described in algorithm 1.

Algorithm 1: K-Means Clustering Method

Input: testing dataset, keys, and positions
Output: group the records either to normal cluster or attacks cluster

1. Search for keys based on its positions (Misuse Method):

attack_label ← 0
for i ← 1 to total number of keys (5 keys)
for j ← 1 to the total number of positions
if Block (i, position (i, j)) = key (i, position (i, j))
attack_label ← 1
end if
end for
end for

2. Search for keys based on its positions (Anomaly Method):

normal_label ← 0
for i ← 1 to total number of keys (6 keys)
if Block (i, position (j)) = key (i, position (i))
normal_label ← 1
end if
end for

3. check the results:

If attack_label = 0 **and** normal_label = 0
record is miss-classified
else if attack_label = 1 **and** normal_label = 0
record is within attack cluster
else if attack_label = 0 **and** normal_label = 1
record is within normal cluster
else if attack_label = 1 **and** normal_label = 1
record is multi-classified
end if
end if
end if
end if

As shown in the aforementioned algorithm 1, the clustering method is done based on three steps. Based on the misuse method, the first step involves the movement of a key (attack signature keys) to a specific position of record which equal to key positions and check if the key is equal to the record block, then put the attack label as 1. The second step, based on the anomaly method, moves the key (user behavior keys) to a specific position of record equal to key positions and checks if the key is equal to record block, then put a normal label as 1. In the final step, check the results, where four different possible cases are obtained:

1. If Attack label =1 and Normal label =0, then the record is classified as Attack (this means the record contains a block that equal to the attack's key only).
2. If Attack label =0 and Normal label =1, then the record is classified as Normal (this means the record contains a block that equal to normal's key only).
3. If Attack label =1 and Normal label =1, then the record is classified as Multi-classified (this means the record contains blocks that equal to both attack's key and normal's key).
4. If Attack label =0 and Normal label =0, then the record is classified as Miss-classified (this means the record does not contain blocks that equal either attack's key or normal's key).

Results and Discussion

The UNSW-NB15 dataset is utilized in the currently proposed system for both training and testing phases as one of the newly available data sets used to evaluate the IDS, and this dataset consists of 10 different records types (normal records and nine different attack types records). The present work involves two main steps, in the first place, the training dataset is used to extract attacks keys and positions, and also to extract normal keys and positions. While in the second step, part of the testing dataset (4000 random records) was used to evaluate the proposed system. The performance of the proposed system is determined by five measures, namely: detection rate (DR), false alarm rate (FAR), accuracy, encoding time, and the time needed for clustering.

The clustering results based on the misuse method (depending on attack records) are shown in Figure 1, while clustering results based on the anomaly method (depending on normal records) is shown in Figure 2. Finally, the clustering results based on a hybrid method (depending on both normal and attack records) are shown in Table 8 and are presented in Figure 3.

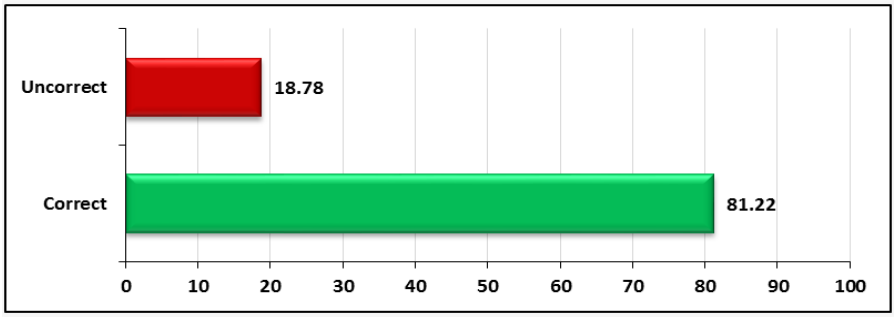

Figure 1 The achieved detection rate results based on the misuse method

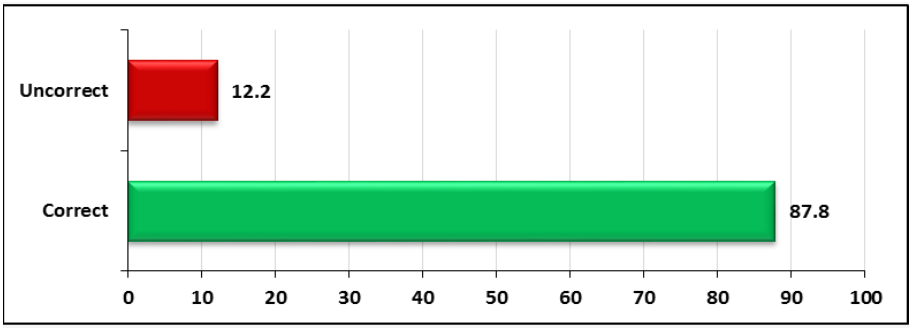

Figure 2 The achieved detection rate results based on the anomaly method

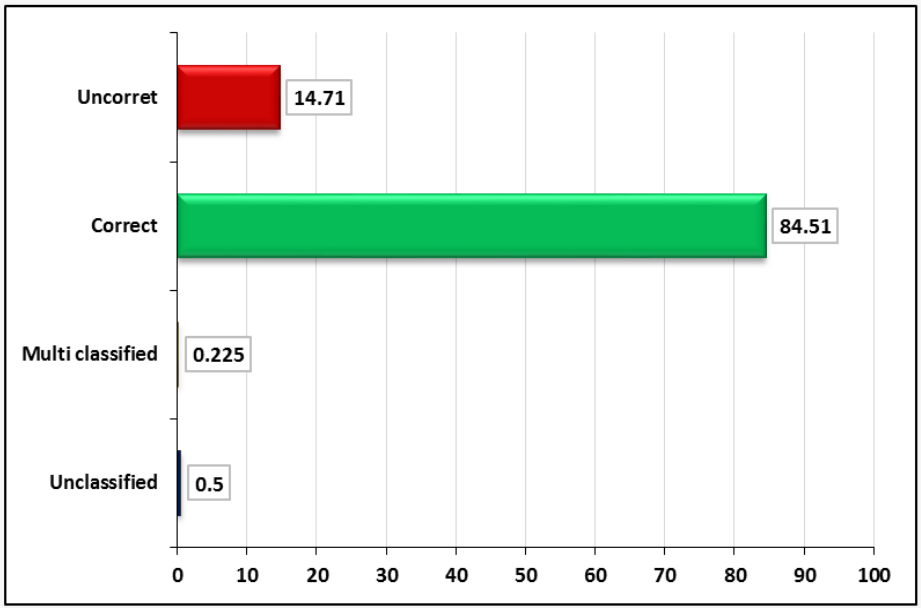

Figure 3 The achieved detection rate results based on a hybrid method

Table 8 The achieved detection rate results based on a hybrid method

Records	**Clustering results**
Attack records	81.22%
Normal records	87.8%
Miss-classified records	1%
Multi-classified records	0.55%

As outlined in Table 8 and Figure 3, results for the correct clustering of attack records is equal to 81.22%, for the correct clustering of normal records is equal to 87.8%, for the total unclassified records are equal to 1%, and for the total multiple classified records is equal to 0.55%.

The achieved DR, FAR, and accuracy results are shown in Table 9. From the table, the DR value is equal to 81.22%, the FAR value is equal to 12.2%, and the accuracy value is equal to 82.05%.

Table 9 DR, FAR, and accuracy results

	Result (%)
DR	81.22
FAR	12.2
Accuracy	82.05

Table 10 and Figure 4 show the times required for encoding and clustering processes. From this table, the encoding and clustering times are equal to 0.385 seconds and 0.00325 seconds for one record respectively. This means that the obtained times by the proposed method are very fast.

Table 10 Encoding and clustering times

	Time (second)
Total encoding time for all 4000 records	1541
Encoding time for one record	0.385
Total clustering time for all 4000 records	13
Clustering time for one record	0.00325

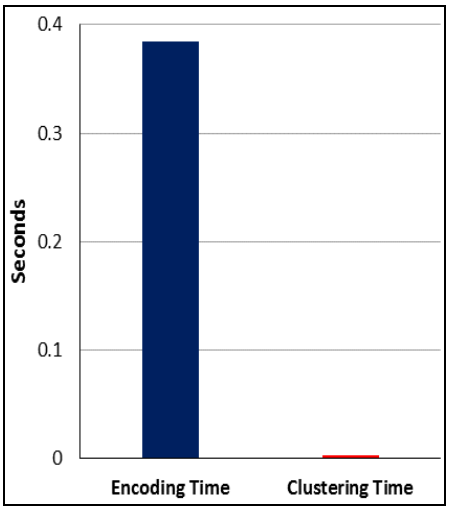

Figure 4 The calculated encoding time and matching time for one record by applying the hybrid method

The comparison between the outcomes of the proposed method and two published IDS (Mebawondu et al. (Mebawondu et al, 2020) and Jing and Chen (Jing et al, 2019)), where the comparison is done in terms of DR, FAR and accuracy, and all papers that we used for comparison depend on UNSW-NB15 dataset. From the table, it is clear that the DR, FAR, and accuracy obtained by the proposed approach are good.

Table 11 Comparison of DR, FAR and accuracy of proposed methods and state-of-the-art

	DR	**FAR**	**Accuracy**
Proposed Method	81.22	12.2	82.05
Mebawondu et al. (Mebawondu et al, 2020)	77	20	76.96
Jing and Chen (Jing et al, 2019)	-	15.26	85.99

Conclusions

A hybrid intrusion detection system is proposed based on the combination of a DNA encoding method and the clustering method. Within this context, a new DNA encoding method is proposed based on the UNSW-NB15 data set by dividing the record's attributes into four groups, namely State, Protocol, Service, and the rest of the features are Digits. A good detection rate and accuracy values are achieved by the proposed approach that equals 81.22% and 82.05% respectively. Also, the encoding time and clustering time is very fast and are equal to 0.385 second and 0.00325 seconds respectively for one record. The achieved good results highlight the importance of the present hybrid IDS to be followed in future in such field of intrusion detection methods.

References

Abdurrazaq, M.N., Bambang, R.T., & Rahardjo, B. (2014). Distributed intrusion detection system using cooperative agent based on ant colony clustering. *In International Conference on Electrical Engineering and Computer Science (ICEECS)*, 109-114.

Aissa, N.B., & Guerroumi, M. (2015). A genetic clustering technique for Anomaly-based Intrusion Detection Systems. *In IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 1-6.

Al-Ibaisi, T., Abu Dalhoum, A. L., Al-Rawi, M., Alfonseca, M., & Ortega, A. (2008). Network intrusion detection using genetic algorithm to find best DNA signature. *WSEAS Transactions on Systems.*

Bharti, K.K., Shukla, S., & Jain, S. (2010). Intrusion detection using clustering. *International Journal of Computer and Communication Technology, 1*(4), 248-255.

Chahal, J.K., & Kaur, A. (2016). A hybrid approach based on classification and clustering for intrusion detection system. *International Journal of Mathematical Sciences & Computing, 2*(4), 34-40.

Chen, Y. (2019). Research on Multi-layer Adaptive Intrusion Detection Based on Clustering and Neural Network. *In 14th International Conference on Computer Science & Education (ICCSE)*, 1-4.

Choudhary, S., & Kesswani, N. (2019). Cluster-based intrusion detection method for internet of things. *In IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, 1-8.

Chung, Y.Y., & Wahid, N. (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO). *Applied soft computing, 12*(9), 3014-3022.

Desai, A.S., & Gaikwad, D.P. (2016). Real time hybrid intrusion detection system using signature matching algorithm and fuzzy-GA. *In IEEE international conference on advances in electronics, communication and computer technology (ICAECCT)*, 291-294.

Eslamnezhad, M., & Varjani, A.Y. (2014). Intrusion detection based on MinMax K-means clustering. *In 7th International Symposium on Telecommunications (IST'2014)*, 804-808.

Hameed, S.M., & Rashid, O.F. (2014). Intrusion detection approach based on DNA signature. *Iraqi Journal of Science, 55*(1), 241-250.

Horng, S.J., Su, M.Y., Chen, Y.H., Kao, T.W., Chen, R.J., Lai, J.L., & Perkasa, C.D. (2011). A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert systems with Applications, 38*(1), 306-313.

Jing, D., & Chen, H.B. (2019). SVM based network intrusion detection for the UNSW-NB15 dataset. *In IEEE 13th International Conference on ASIC (ASICON)*, 1-4.

Liang, D., Liu, Q., Zhao, B., Zhu, Z., & Liu, D. (2019). A clustering-svm ensemble method for intrusion detection system. *In 8th International Symposium on Next Generation Electronics (ISNE)*, 1-3.

Liang, W., Li, K. C., Long, J., Kui, X., & Zomaya, A.Y. (2019). An industrial network intrusion detection algorithm based on multifeature data clustering optimization model. *IEEE Transactions on Industrial Informatics, 16*(3), 2063-2071.

Mebawondu, J.O., Alowolodu, O.D., Mebawondu, J.O., & Adetunmbi, A.O. (2020). Network intrusion detection system using supervised learning paradigm. *Scientific African, 9.*

Modi, C.N., & Patel, D. (2013). A novel hybrid-network intrusion detection system (H-NIDS) in cloud computing. *In IEEE Symposium on Computational Intelligence in Cyber Security (CICS)*, 23-30

Moustafa, N., & Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *In military communications and information systems conference (MilCIS)*, 1-6.

Mukherjee, S., & Sharma, N. (2012). Intrusion detection using naive Bayes classifier with feature reduction. *Procedia Technology, 4*, 119-128.

Roshan, S., Miche, Y., Akusok, A., & Lendasse, A. (2018). Adaptive and online network intrusion detection system using clustering and extreme learning machines. *Journal of the Franklin Institute, 355*(4), 1752-1779.

Rustam, Z., & Maharani, J. (2020). Intrusion detection system model using Fuzzy Kernel C-Means and Laplacian Score feature selection. *In Journal of Physics: Conference Series*, 1442(1).

Rustam, Z., & Talita, A.S. (2018). October. Fuzzy kernel robust clustering for anomaly based intrusion detection. *In Third International Conference on Informatics and Computing (ICIC)*, 1-4.

Shojafar, M., Taheri, R., Pooranian, Z., Javidan, R., Miri, A., & Jararweh, Y. (2019). Automatic clustering of attacks in intrusion detection systems. *In IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, 1-8.

Sumanth, R., & Bhanu, K.N. (2020). Raspberry Pi Based Intrusion Detection System Using K-Means Clustering Algorithm. *In Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 221-229.

Verma, P., Anwar, S., Khan, S., & Mane, S.B. (2018). Network intrusion detection using clustering and gradient boosting. *In 9th International conference on computing, communication and networking technologies (ICCCNT)*, 1-7.

Wang, C., Huang, R., Zhang, W., & Sun, J. (2019). Multilayer Intrusion Detection System Based On Semi-supervised Clustering. *In 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, 355-360.

Wang, S. (2011). Research of intrusion detection based on an improved K-means algorithm. *In Second International Conference on Innovations in Bio-inspired Computing and Applications*, 274-276.

Wei, L., Zhong-Ming, Y., Ya-Ping, C., & Bin, Z. (2017). A clustering algorithm oriented to intrusion detection. *In IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 1, 862-865.

Biographies of Authors

	**Omar Fitian Rashid** was born in Baghdad, Iraq in 1988. He got his BSc in computer science and MSc in computer security from the College of Science, University of Baghdad in 2010 and 2014. He has successfully defended his PhD thesis in 2019 entitled 'Enhanced DNA Encoding for Anomaly Intrusion Detection System' at the Faculty of Information Science and Technology, National University of Malaysia (Universiti Kebangsaan Malaysia). His main field of interest is the computer and network security.
	**Mazin Sameer Al-Hkeem** was born in Iraq in 1978. He got his BSc and MSc in computer science and PhD in computer science from University of Technology in 2007. He has as 13 years' experience on Network Technology & Security and Web Technology. Published a few scientific books and several researches in many international conferences and scientific journals. His main field of interest is the IoT and network security.