Denial of Service Attacks Detection on SCADA Network IEC 60870-5-104 using Machine Learning

By M. Agus Arifin

WORD COUNT 3137
TIME SUBMITTED 02-AUG-2021 09:50AM
PAPER ID 75130420

Denial of Service Attacks Detection on SCADA Network IEC 60870-5-104 using Machine Learning

M. Agus Syamsul Arifin
Faculty of Computer Universitas Bina Insan/Faculty of Engineering Universitas Sriwijaya
Lubuklinggau, Indonesia
mas.arifin@univbinainsan.ac.id

Deris Stiawan*
Computer Engineering Department, Faculty of Computer Science Universitas Sriwijaya
Palembang, Indonesia
deris@unsri.ac.id

Susanto
Faculty of Computer Universitas Bina Insan/Faculty of Engineering Universitas Sriwijaya
Lubuklinggau, Indonesia
susanto@univbinainsan.ac.id

Juli Rejito
Department of Computer Science, Faculty of Math. and Natural Sciences Universitas Padjadjaran
juli.rejito@unpad.ac.id

Mohd. Yazid Idris
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia
yazid@utm.my

Rahmat Budiarto
College of Computer Science and IT, AlBaha University
Albaha, Saudi Arabia
rahmat@bu.edu.sa

***Abstract*— SCADA was designed to be used in an isolated area however, in modern SCADA, its connection to the Internet has become essential due to performance and commercial needs. This extended SCADA interconnection creates new vulnerabilities in SCADA network. One of the attacks that may occur caused by extended interconnection of SCADA networks to heterogeneous networks is Denial of Service attacks (DoS). DoS attack is launched by sending many messages from nodes. The development of easily accessible and simple DoS tools has increased the frequency of attacks. Ease of access and use of DoS tools made reduced the level of expertise needed to launch an attack. This study uses a SCADA dataset containing DoS attacks and running IEC 60870-5-104 protocol where this protocol will be encapsulated into TCP/IP protocol before being transmitted so that the treatment in detecting DoS attack in SCADA networks using the IEC 104 protocol is not much different from a traditional computer network. This study implements three machine learning approaches, i.e.: Decision Tree, Support Vector Machine, and Gaussian Naïve Bayes in creating an Intrusion Detection System (IDS) model to recognize DoS attack on the SCADA Network. Experimental results show that the performance of the Decision Tree approach has the best performance detection on the Testing dataset and Training dataset with an accuracy of 99.99% in all experiments.**

Keywords—SCADA, IEC 60870-5-104, Denial of Service, Intrusion Detection System

I. INTRODUCTION

The legacy of the SCADA systems is operated and designed for isolated networks, which make SCADA systems less exposed from the Internet network so made them receives fewer threats from internet network [1][2][3]. Connecting critical infrastructures to the Internet has become essential due to performance and commercial needs [4]. The interconnection of the SCADA network to the corporate network allows business users to access real-time data generated by SCADA, at the same time it opens security holes in the SCADA system [2]. Due to the interconnection, the potential for cyberattacks increases drastically [5]. Currently, Denial of Service attacks is the biggest threat in the internet network [6], development of easily accessible and simple DoS tools has increased the frequency of attacks. Ease of access and use of DoS tools made reduced the level of expertise needed to launch an attack [7].

DoS (Denial of Service) attack is launched by sending many messages from nodes/computers, in order to overwhelm the company's servers and paralyze its website for several hours, to block access to Internet users [8]. This scenario also happens to the SCADA network infrastructure which is increasingly vulnerable to DoS attacks due to the interconnection of SCADA to heterogeneous networks, especially SCADA systems used by power plants. In this study, the authors use a SCADA dataset with the IEC 60870-5-104 (IEC 104) protocol where the IEC 104 protocol will be encapsulated into TCP IP protocol before being transmitted [9][10][11]. The IEC 104 protocol is widely used because it can use Automatic Generation Control (AGC) where the algorithm can adjust the electric power balance on a wide geographic scale [12].

This study discusses the performance of machine learning classification algorithms such as Decision Tree, Support Vector Machine, and Gaussian Naïve Bayes in detecting DoS attacks in SCADA IEC 104 networks. Precision, Recall, F-Measure, and Accuracy of each algorithm are compared to investigate the best algorithm. The dataset created by M. Egger [13] is used. The dataset contains several malicious activities, however, in this study, the authors only focus on how to detect DoS activity using machine learning algorithms. This work contributes towards the development of IDSs for SCADA systems, especially for SCADA running IEC 60870-5-104 protocol.

II. RELATED WORK

Several studies on DoS attack detection have been carried out on traditional computer networks that use TCP/IP protocol. SCADA network with IEC 104 protocol usually uses TCP/IP protocol for transmitting a data packet [9][10] [11], so that DoS attacks on traditional computer networks can also be launched on SCADA networks running the IEC 104 protocol. Therefore, to detect DoS attack activities on SCADA networks running IEC 104 protocol have similarities with DoS attacks detection on traditional computer networks but with adjustments for data

packets that contain IEC 104, namely APCI (Application Protocol Control Information) and ASDU (Application Service Data Unit).

According to [14] the common practice and order of the message (codification) in the IEC 104 protocol widely used in the power control domain is called Network Time Protocol (NTP) [14]. However, the NTP mechanism can be weakness in the IEC 104 as research conducted by [14] describe attack patterns allowing the undetected partial replay of valid messages and injection of messages measures in a multi-staged attack targeting the NTP protocol, for resulting in a de-synchronisation between a PLC (Programmable Logic Controller) device and/or RTU (Remote Terminal Unit) device and higher SCADA components [14] thus disrupting on the SCADA network communication. This activity is evolved as a DoS attack and makes Remote Terminal Unit (RTU) crashes or blocks the other message which goes to the RTU. In general, the devices that make up the SCADA network, especially the RTU has low computed resources [15], so that if a DoS attack occurs it will greatly affect the performance of the device and the entire SCADA system.

Research conducted by [6] uses machine learning and neural network algorithms to detect DoS attacks using the 2017 CICDS dataset, then research performed by [16] uses intelligent evolutionary algorithms to defend against DoS attacks. The research was done by [17] uses software-defined networking in defence of DoS attacks and port scans for efficiency and defences in real-time. In a study conducted by [7], the authors use data mining to mitigate DoS attacks on computer networks with self-generated datasets. DoS attack detection research works were also carried out by [18] and [19] in building an intrusion detection system using machine learning on a smart grid to detect DoS attacks. Then research on SCADA for the problem of DDoS attacks was carried out by [20] to determine the pattern of DDoS attacks using also machine learning. Furthermore, according to [21], traditional intrusion defence for IT systems often cannot be applied on substation SCADA networks running IEC 104 as described in their research in detecting DoS and Man in the Middle (MITM) attacks on SCADA networks.

III. Design and Method

A. Denial-of-Service on Dataset

This study uses the supervised learning method, to simulate a DoS attack against the testing substation, the application "hping" was used with the command: *hping3 -flood –S 192.168.0/24*. the -flood parameter is used in the program to send packets as fast as possible while -S symbolized the SYN Flood attack [13].

During the DoS attack, *hping* command tries to send as many as possible SYN requests to a device in the testing substation which caused them to crash. In addition to the SYN packets, the next feature is the time to live (TTL) which is lower than 64 as a DoS feature on the dataset used because a TTL of 64 is a feature of port scan activity in this dataset [13].

B. Proposed Model

In the dataset, there are several types of attacks, however, in this study, the authors only focus on detecting DoS attacks whose attacks are aimed at the RTU devices on the testbed network. The dataset in this study uses the CSV format, cleaning and normalization of the dataset are performed first and then only needed data are used. Figure 1 shows the workflow to determine the best machine learning algorithm to be used for the IDS on the SCADA network.

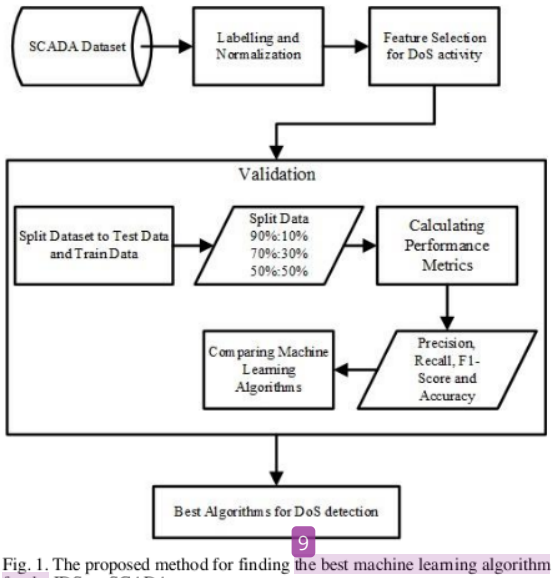

Fig. 1. The proposed method for finding the best machine learning algorithm for the IDS on SCADA.

At the validation stage in choosing the best machine learning algorithm, the authors conducted three experiments in the first experiment, dividing the dataset into 90% training data and 10% testing data; the second experiment, dividing the dataset into 70% training data and 30% testing data; then on the third experiment, dividing the dataset into 50% training data and 50% testing data. Dividing of the dataset to provide accurate validation data for each measurement of the machine learning algorithm used in this study.

C. Classifier Algorithm

We used three classifier algorithms of machine learning for IDS, i.e.: Decision Tree (DT), Gaussian Naïve Bayes (NB) and Support Vector Machine (SVM). The three machine learning algorithms are categorized into supervised learning algorithm.

The supervised algorithm deals with fully class labelled data and finds the relationship between data and its class. The classification has two steps, i.e.: training and testing. The training data is done with the help of the response variable [22].

D. Methodology

The IDS model implements machine learning algorithms to detect DoS attacks and is generated from data training, the dataset is used as a learning medium for the IDS model then the performance of the model is measured. Metrics for

evaluating the model performance include True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values. Table 1 shows the format of the confusion matrix.

TABLE I. CONFUSION MATRIX

Actual/Predicted Class	Normal	DoS
Normal	TN	FP
DoS	FN	TP

Items from the table can be explained as follows :

- TN : Actual Normal Data is classified as Normal Data
- FP : Actual Normal Data is classified as DoS Attack
- FN : Actual DoS Attack is classified as DoS Attack
- TP : Actual DoS Attack is classified as Normal Data

Table 2 describes the measurement of Accuracy, Recall, Precision, and F1-Score (F-Measure) performance metrics which are calculated using instances from the Confusion Matrix in Table 1.

TABLE II. PERFORMANCE MATRIX

Measure	Formula
Accuracy	(TP + TN) / (TP+FP+FN+TN)
Recall	TP / (TP+FN)
Precision	TP / (TP+FP)
F1-Score	2TP / (2TP+FP+FN)

IV. RESULT

The evaluation process for this experiment is run using CPU, it running on a Ubuntu 20.04 operating system with Intel(R) Core(TM) i7-9750H and 32Gb Ram. The use of GPU will be considered in the future for reducing training time.

The number of records in the dataset after the normalization process was 248,278 records with 245,777 normal and 2501 DoS attack activity. In this experiment, the authors do not use the oversampling technique to balance the DoS class with the normal class because the accuracy and other performance measurements such as precision, recall and F1-Score obtained in the experiments show good results. The metrics performance measurement of each classification algorithm on the testing dataset and training dataset are shown in Table 3 for accuracy performance, Table 4 for Recall, Precision and F1-Score performance.

TABLE III. COMPARISON OF ACCURACY ON DATA TEST AND DATA TRAIN

Data Split	Data of Class	Accuracy (%)			No. Class
		DT	*GNB*	*SVM*	
Data Test	10 %	99.99	95.51	99.97	24828
	30 %	99.99	95.49	99.97	74484
	50 %	99.99	95.52	99.98	124139
Data Train	50 %	99.99	95.52	99.97	124139
	70 %	99.99	95.53	99.97	173794
	90 %	99.99	95.52	99.97	223450

From experimental results the accuracy of the three algorithms are above 90%, however, the accuracy obtained does not indicate the number of false alarms of the IDS. The high accuracy of this study is also caused by the number of DoS classes is only 1% of the normal data class, especially on the accuracy results of the Gaussian Naïve Bayes algorithm does not describe the good performance of the built IDS model with this algorithm even though the accuracy obtained is more than 95%. Therefore Precision, Recall, F1-Score and Confusion Matrix measurements are a way to validate the performance of the IDS model. More detailed observations on the performance of the machine learning algorithms can be seen in Table 4, Figure 2 and Figure 3 for the Confusion matrix of each experiment.

TABLE IV. COMPARISON OF CLASS PERFORMANCE MATRIX ON THE TESTING DATASET AND TRAINING DATASET

Data Split	Data of Class	Class \ Measure	Precision (%)			Recall (%)			F1-Score			No. Class
			DT	*GNB*	*SVM*	*DT*	*GNB*	*SVM*	*DT*	*GNB*	*SVM*	
Data Test	10 %	Normal	100	95	100	100	100	100	100	98	100	24828
		DoS	99	100	98	100	17	99	99	30	99	
	30 %	Normal	100	95	100	100	100	100	100	98	100	74484
		DoS	99	100	98	100	18	100	99	30	99	
	50 %	Normal	100	95	100	100	100	100	100	98	100	124139
		DoS	99	100	98	100	18	100	99	30	99	
Data Train	50 %	Normal	100	100	100	100	95	100	100	98	100	124139
		DoS	100	19	99	99	100	98	100	32	99	
	70 %	Normal	100	100	100	100	95	100	100	98	100	173794
		DoS	100	19	99	99	100	98	99	31	99	
	90 %	Normal	100	100	100	100	95	100	100	98	100	223450
		DoS	100	18	99	99	100	99	99	31	99	

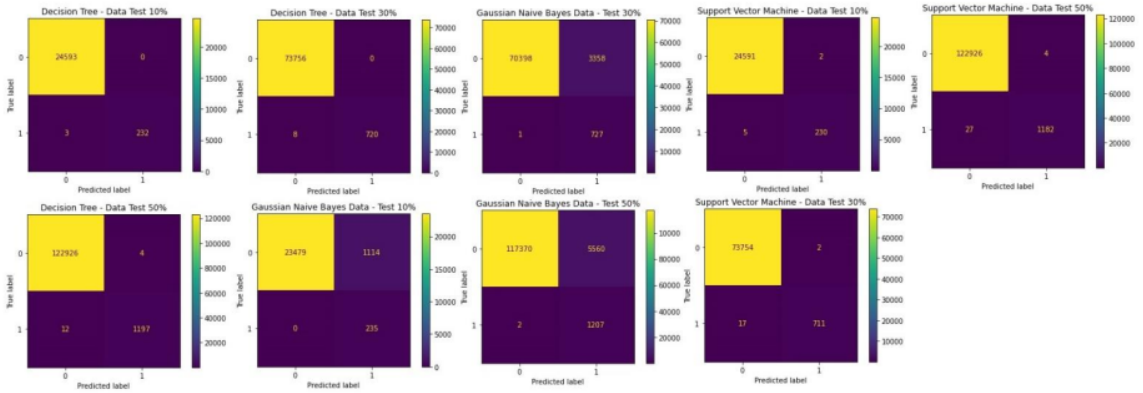

Fig. 2. Comparison of Confusion Matrix on the Data Test

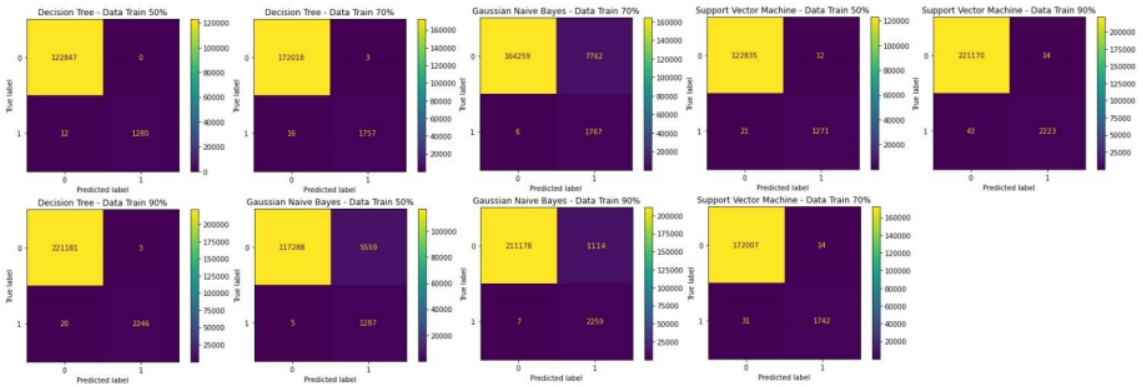

Fig. 3. Comparison of Confusion Matrix on the Data Train

Observations and measurements on the performance metrics show that the Gaussian Naïve Bayes algorithm has a high false alarm rate for each experiment, both for testing data and training data. While the best performance in terms of accuracy, precision, recall and F1-Score is the Decision Tree algorithm which shows the best performance compared to the Support Vector Machine and Gaussian Naïve Bayes algorithms, as depicted in Table 3 and Table 4. The decision tree algorithm also has a low false alarm on every test performed.

V. Conclusion and Future Work

From the experimental results, the performance of the Decision Tree algorithm has the best detection with an accuracy of 99.99% for all experiment scenarios, the decision tree also has a low false alarm on every test performed. Precision, recall and F1-Score performance of Decision Tree algorithm also showed the best performance compared to the Gaussian Naïve Bayes algorithm and the Support Vector Machine algorithm which were also measured to build an intrusion detection system in this study.

For future work, the authors plan to create a dataset from SCADA network testbed that running IEC 60870-5-104 protocol with more complete attack data, especially attacks data that only specifically occurs on the SCADA IEC 60870-5-104 system such as activities from unauthorized devices using TESTFR, STARTDT, and STOPDT data packets on SCADA networks where these packets are only found in the protocols of SCADA system.

References

[1] D. Pliatsios, P. Sarigiannidis, T. Lagkas, and A. G. Sarigiannidis, "A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 3, pp. 1942–1976, 2020.

[2] M. R. Asghar, Q. Hu, and S. Zeadally, "Cybersecurity in industrial control systems: Issues, technologies, and challenges," *Comput. Networks*, vol. 165, 2019.

[3] J. E. Rubio, C. Alcaraz, R. Roman, and J. Lopez, "Current cyber-defense trends in industrial control systems," *Comput. Secur.*, vol. 87, p. 101561, 2019.

[4] M. Kalech, "Cyber-attack detection in SCADA systems using temporal pattern recognition techniques," *Comput. Secur.*, vol. 84, pp. 225–238, 2019.

[5] A. Volkova, M. Niedermeier, R. Basmadjian, and H. De Meer, "Security challenges in control network protocols: A survey," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 1, pp. 619–639, 2019.

[6] S. Wankhede and D. Kshirsagar, "DoS Attack Detection Using Machine Learning and Neural Network," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, 2018.

[7] K. Soni and S. Singh, "A proposed DoS detection scheme for mitigating DoS attack," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 10, no. 4, pp. 172–179, 2020.

[8] S. Tamy, H. Belhadaoui, M. A. Rabbah, N. Rabbah, and M. Rifi, "SCADA Communication Real Time Protocols," *Indian J. Sci. Technol.*, vol. 12, no. 34, pp. 1–6, 2019.

[9] P. Eden, A. Blyth, P. Burnap, Y. Cherdantseva, K. Jones, and H.

Soulsby, “A Forensic Taxonomy of SCADA Systems and Approach to Incident Response,” pp. 42–51, 2015.

[10] W. Hou, X. Zhang, L. Guo, Y. Sun, S. Wang, and Y. Zhang, “Taxonomy of attacks on Industrial Control protocols,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9784, pp. 78–87, 2016.

[11] P. Matoušek, O. Ryšavý, and M. Grégr, “Increasing Visibility of IEC 104 Communication in the Smart Grid,” no. December 2015, pp. 21–30, 2019.

[12] K. Mai, X. Qin, N. Ortiz Silva, and A. A. Cardenas, “IEC 60870-5-104 network characterization of a large-scale operational power grid,” *Proc. - 2019 IEEE Symp. Secur. Priv. Work. SPW 2019*, pp. 236–241, 2019.

[13] M. Egger, G. Eibl, and D. Engel, “Comparison of Approaches for Intrusion Detection in Substations using the IEC 60870-5-104 Protocol,” *Energy Informatics*, vol. 3, no. Suppl 1, p. to appear, 2020.

[14] A. Baiocco and S. D. Wolthusen, “Indirect Synchronisation Vulnerabilities in the IEC 60870-5-104 Standard,” *Proc. - 2018 IEEE PES Innov. Smart Grid Technol. Conf. Eur. ISGT-Europe 2018*, pp. 1–6, 2018.

[15] C. E. Texas, “Bloom Filter Based Intrusion Detection for Smart Grid Scada,” *Saranya Parthasarathy, Deep. Kundur*, no. May, 2012.

[16] S. Dwivedi, M. Vardhan, and S. Tripathi, “Defense against distributed DoS attack detection by using intelligent evolutionary algorithm,” *Int. J. Comput. Appl.*, vol. 0, no. 0, pp. 1–11, 2020.

[17] C. Birkinshaw, E. Rouka, and V. G. Vassilakis, “Implementing an intrusion detection and prevention system using software-defined networking: Defending against port-scanning and denial-of-service attacks,” *J. Netw. Comput. Appl.*, vol. 136, no. March, pp. 71–85, 2019.

[18] Z. Wang, W. Cheng, and C. Li, “DoS attack detection model of smart grid based on machine learning method,” *Proc. 2020 IEEE Int. Conf. Power, Intell. Comput. Syst. ICPICS 2020*, pp. 735–738, 2020.

[19] P. I. Radoglou-Grammatikis and P. G. Sarigiannidis, “Securing the Smart Grid: A Comprehensive Compilation of Intrusion Detection and Prevention Systems,” *IEEE Access*, vol. 7, pp. 46595–46620, 2019.

[20] F. A. Alhaidari and E. M. Al-Dahasi, “New approach to determine DDoS attack patterns on SCADA system using machine learning,” *2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019*, pp. 1–6, 2019.

[21] P. Kreimel, O. Eigner, F. Mercaldo, A. Santone, and P. Tavolato, “Anomaly detection in substation networks,” *J. Inf. Secur. Appl.*, vol. 54, 2020.

[22] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan, “Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review,” *Procedia Comput. Sci.*, vol. 171, no. 2019, pp. 1251–1260, 2020.

Denial of Service Attacks Detection on SCADA Network IEC 60870-5-104 using Machine Learning

ORIGINALITY REPORT

20%
SIMILARITY INDEX

PRIMARY SOURCES

1	M. Agus Syamsul Arifin, Susanto Susanto, Deris Stiawan, Mohd Yazid Idris, Rahmat Budiarto. "The trends of supervisory control and data acquisition security challenges in heterogeneous networks", Indonesian Journal of Electrical Engineering and Computer Science, 2021 Crossref	50 words — 2%
2	Susanto, Deris Stiawan, M. Agus Syamsul Arifin, Mohd. Yazid Idris, Rahmat Budiarto. "IoT Botnet Malware Classification Using Weka Tool and Scikit-learn Machine Learning", 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI), 2020 Crossref	48 words — 2%
3	Michael Egger, Günther Eibl, Dominik Engel. "Comparison of approaches for intrusion detection in substations using the IEC 60870-5-104 protocol", Energy Informatics, 2020 Crossref	43 words — 2%
4	export.arxiv.org Internet	40 words — 2%
5	ieeexplore.ieee.org Internet	40 words — 2%

6	T. Saranya, S. Sridevi, C. Deisy, Tran Duc Chung, M.K.A.Ahamed Khan. "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review", Procedia Computer Science, 2020 Crossref	38 words — 1%
7	Meir Kalech. "Cyber-attack detection in SCADA systems using temporal pattern recognition techniques", Computers & Security, 2019 Crossref	26 words — 1%
8	Sharipuddin, Benni Purnama, Kurniabudi, Eko Arip Winanto et al. "Features Extraction on IoT Intrusion Detection System Using Principal Components Analysis (PCA)", 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI), 2020 Crossref	26 words — 1%
9	Natalia V Petrova. BMC Bioinformatics, 2006 Crossref	22 words — 1%
10	Kolenc, Mitja, Nermin Suljanovic, Peter Nemcek, and Matej Zajc. "Monitoring communication QoS parameters of distributed energy resources", 2016 IEEE International Energy Conference (ENERGYCON), 2016. Crossref	16 words — 1%
11	Deris Stiawan, Ahmad Heryanto, Ali Bardadi, Dian Palupi Rini et al. "An Approach for Optimizing Ensemble Intrusion Detection Systems", IEEE Access, 2021 Crossref	13 words — 1%
12	bioone.org Internet	12 words — < 1%
13	www.mdpi.com Internet	12 words — < 1%

14	energyinformatics.springeropen.com Internet	11 words — < 1%
15	Fariz Andri Bakhtiar, Eko Sakti Pramukantoro, Hilman Nihri. "A Lightweight IDS Based on J48 Algorithm for Detecting DoS Attacks on IoT Middleware", 2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech), 2019 Crossref	10 words — < 1%
16	upcommons.upc.edu Internet	10 words — < 1%
17	Soumava Dey, Gunther Correia Bacellar, Mallikarjuna Basappa Chandrappa, Raj Kulkarni. "COVID-19 Chest X-Ray Image Classification Using Deep Learning", Cold Spring Harbor Laboratory, 2021 Crossref Posted Content	9 words — < 1%
18	downloads.hindawi.com Internet	9 words — < 1%
19	"Machine Intelligence and Soft Computing", Springer Science and Business Media LLC, 2021 Crossref	8 words — < 1%
20	Arun Kumar Bediya, Rajendra Kumar. "A Novel Intrusion Detection System for Internet of Things Network Security", Journal of Information Technology Research, 2021 Crossref	8 words — < 1%
21	Dimitrios Pliatsios, Panagiotis Sarigiannidis, Thomas Lagkas, Antonios G. Sarigiannidis. "A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics", IEEE Communications Surveys & Tutorials, 2020 Crossref	8 words — < 1%

22	Lecture Notes in Computer Science, 2016. Crossref	8 words — < 1%
23	B o Mohammed. "Improvement in twins handwriting identification with invariants discretization", EURASIP Journal on Advances in Signal Processing, 2012 Crossref	7 words — < 1%
24	Jackson Kamiri, Geoffrey Mariga. "Research Methods in Machine Learning: A Content Analysis", International Journal of Computer and Information Technology(2279-0764), 2021 Crossref	7 words — < 1%
25	Mohammed. "Feature Discretization for Individuality Representation in Twins Handwritten Identification", Journal of Computer Science, 2011 Crossref	7 words — < 1%
26	"Critical Infrastructure Protection XIII", Springer Science and Business Media LLC, 2019 Crossref	6 words — < 1%
27	"Intelligent Data Communication Technologies and Internet of Things", Springer Science and Business Media LLC, 2020 Crossref	6 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF