Segmentation of Illicit Behaviour in IoT via Artificial Immune Systems

Mikel Moreno Moreno
Fundación Vicomtech
Mikeletegi 57, 20009 San Sebastián
mmoreno@vicomtech.org

Lander Segurola Gil
Fundación Vicomtech
Mikeletegi 57, 20009 San Sebastián
lsegurola@vicomtech.org

Raul Orduna Urrutia
Fundación Vicomtech
Mikeletegi 57, 20009 San Sebastián
rorduna@vicomtech.org

***Abstract*—In recent years due to the increasing number of devices connected to the Internet in what is known as the era of the Internet of Things, the number of potential vulnerabilities has also increased. Various anomaly detectors and malicious behaviour classification algorithms have been proposed. Still, in unsupervised training scenarios, the artificial intelligence models focus on detecting anomalies and do not differentiate between different behaviour patterns. To improve the level of detail for these systems (be able to define entities and group events/messages into homogeneous behaviours) the application of optimization mechanisms based on artificial immune systems (aiNet) in clustering algorithms is proposed.**

The proposed pipeline is comprised of artificial immune systems (aiNet) for generating a first set of detectors, a distance-based clustering method (K-means) for redistributing these detectors and a density-based clustering method (DBSCAN or OPTICS) for refining this clustering and enabling behavioural segmentation.

The system is parametrizable to adapt to the requirements of the search being carried out. In addition, the use of public databases has been made to develop the behaviour extraction model and validate the results with the algorithms for the classification of malicious behaviours and entity identification already available.

***Index Terms*—Cybersecurity, Multi-Label Classification, Immune Network, Clustering Algorithms, Network traffic, Unsupervised Learning**

Contribution Type: *Research in development (limit 8 pages)*

I. Introduction

In recent years, due to the increase in the number of smart devices connected to the internet [1], [2], network security has quickly become an issue of societal concern. The IoT platform generates a large volume of valuable data, which if not securely transmitted and analyzed can lead to a critical privacy breach. Traditional protection mechanisms such as encryption, authentication, and access control are difficult to manage for large systems with multiple connected devices because each part of the system has different inherent vulnerabilities [3], [4]. Consequently, security is at greater risk in IoT systems than in other IT systems, and the traditional solution may be ineffective.

Anomaly, intrusion and cyber attacks traffic identification models using Machine Learning (ML) algorithms for IoT security analysis were proven effective for detecting intrusions that have already been encountered and characterized [5], [6]. However, new unknown threats (often referred to as zero-day attacks or zero-days [7]) likely go undetected as they are often misclassified by those techniques [8].

Unsupervised anomaly detection algorithms do not use labelled information and show the potential to detect zero-days [9]. However, it is acknowledged that unsupervised anomaly detection algorithms may show poor detection performance when used as the sole or main instrument for intrusion detection [10]. In particular, they are likely to generate a high amount of False Positives (the detector raises a security alert but no attacks are happening) and False Negatives (attacks going undetected), thus lowering correct classifications as True Positives or True Negatives.

To solve this problem and offer an unsupervised model that offers a reliable response without making use of labelled information. A system involving optimization mechanisms based on artificial immune systems is proposed.

The rest of the paper is organized as follows: Section 2 provides an overview of the technology used and the related literature; Section 3 details the presented proposal; Section 4 describes the experimental framework introducing the dataset used, feature selection and validation metrics; Section 5 presents and discusses our results; and Section 6 concludes with a summary and suggestions for future work.

II. Background

In this section, the concept of Artificial Immune Networks and the clustering algorithms used in this project are introduced. Also, a brief overview of the evolutionary algorithms for anomaly detection in cybersecurity is discussed.

The *artificial immune network* (aiNet) is part of the artificial immune system theories that are inspired by their biological counterparts and possess similar attributes (self-learning, self-adaptation, self-organization and immune memory) [11], [12]. The aiNet model can be used to refine some important characters of complex information data. At present, its application areas mainly include data clustering, pattern recognition data compression, etc. The goal of it is to find the optimal memory antibodies for each antigen Ag_j using immune evolution strategies (see figure 1).

In biology, antibodies are generated for antigens, i.e. when a known antigen enters the body it is identified and neutralized by previously generated antibodies. In the computational version identifying antigens is equivalent to identifying a set of points, i.e. modeling a behavior, where:

The antigen represents a subset of training samples used to generate the prediction model. In this case, it is the instances of the same attack that are passed as training to the prediction model. The antibody denotes a candidate solution of the prediction model used to approximate antigen behaviour.

ISBN:978-84-09-62140-8

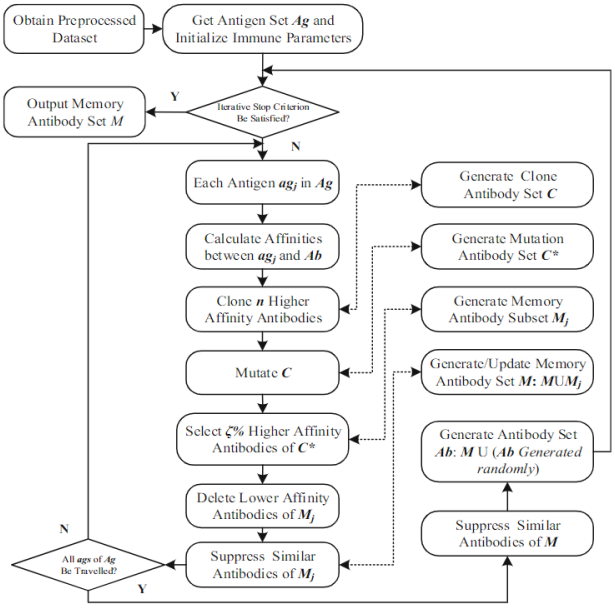

Figure 1. The flowchart of aiNet model [13]

Antigen affinity is defined as the measure for choosing the best set of antibodies for each antigen. It determines how well an antigen is being represented by the antibodies generated.

The K-means is an algorithm aimed at discovering k clusters in the data by applying iterative refinement over a predefined number k of centroids to generate a final result [14], [15]. In other words, for a set of data points, it finds a distribution of k nonempty clusters covering the whole dataset. The main limitations of K-means clustering are the selection of the optimal hyperparameter K and the lack of robustness, due to the instability of the obtained results depending on the initialization of centroids.

Density-based clustering (DBC) refers to unsupervised learning methods that identify distinctive groups or clusters in data based on the idea that a cluster in a data space is a contiguous region of high point density, separated from other such clusters by contiguous regions of low point density. Data points in the low point density separation regions are often considered noise or observations.

The DBSCAN algorithm works by grouping points considered nearby neighbours for a minimum amount of them [16]. In other words, it identifies dense data point regions to cluster them. It has two basic hyperparameters to take into account: ε (specifies how close points must be to each other to be considered as nearby neighbours) and `MinPts` (describes the minimum number of points to form a dense region formed by nearby neighbours).

The OPTICS algorithm works similarly to DBSCAN (requiring the same two parameters) but instead of assigning cluster memberships, it stores the order in which the points are processed [17]. Thus, the points of the database are linearly ordered so that spatially closest points become neighbours in the ordering. Additionally, a special distance is stored for each point that represents the density that must be accepted for a cluster so that both points belong to the same cluster like the latter algorithm.

A. Related work

On the one hand, K-means algorithms can be applied to anomaly detection by distinguishing normal from abnormal behaviour using feature similarity calculations [18], [19]. However, Laskov et al. [20] indicate that while unsupervised algorithms are often a good choice when it is difficult to generate the labelled data, their performance including that of K-means is less effective than that of supervised learning methods in detecting known attacks. This makes it clear that the application of clustering methods, in general, and K-means, in particular, to the security of IoT systems is still in its infancy and needs to be further explored. Therefore, in this work, their use is proposed to obtain new results that reflect the usefulness of these methods in the field of pattern identification.

On the other hand, the use of evolutionary systems for the optimization of segmentation processes goes back to the origins of evolutionary methods, one of which is the artificial immune network (aiNet)[11]. At present, their application areas mainly include data clustering, pattern recognition and data compression [21], [11], [22].

In recent years, artificial immune networks have been employed by intrusion detection systems to cluster anomalous malicious behaviours. Liu et al. [6] proposed an unsupervised anomaly detection algorithm based on an artificial immune network, and agglomerative hierarchical clustering is employed to assist the clustering analysis.

Lau et al. [23] proposed an unsupervised anomaly detection architecture that is capable of online adaptation inspired by immune network theory. Rassam et al. [24] investigated the artificial immune network to cluster the malicious attacks of the intrusion detection system, and employed the rough set principle to obtain the features of the key elements of the given dataset to improve the detection rate of this system. These above anomaly detection approaches demonstrate that the artificial immune network can be effectively used to cluster the network flows and refine the detectors of the anomaly detection system.

Shi et al. [25] proposed the *immunity-based time series prediction approach for network security situation* (*ITSPA*) to effectively improve the accuracy of network security situation prediction and prevent large-scale network security attacks, immunity-based time series prediction approach for network security situation. After that, Shi et al. [13] proposed an unsupervised anomaly detection approach for network flow using Immune Network based K-means clustering (*Unsupervised Anomaly Detection approach for network flow using Immune Network based K-means clustering* or *UADINK*). In this paper, aiNet_KMC is introduced to cluster the network flow by combining the Artificial Immune Networks (aiNet) to analyze and filter the raw dataset to construct an internal image of all data points (a refined relationship map) using immune evolution mechanisms. Therefore, artificial immune networks can be used to refine some important characteristics of complex information data, the optimal value of k for K-means clustering and the aforementioned method to generate the clusters. Next, a cluster labelling algorithm (clusLA) is used to determine whether a cluster is malicious or not.

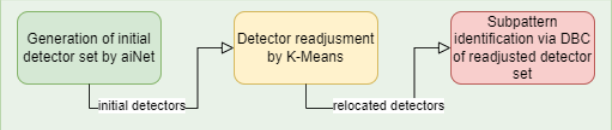

Figure 2. Flowchart of the proposed approach

Finally, the labelled clusters are considered detectors to identify anomalous network flows using the NFAD algorithm (*network flow anomaly detection algorithm*). Subsequently, Shi et al. [26] continued their research with ADAID (*Anomaly Detection approach for network flow using Artificial Immune network and Density peak*), where instead of a combination of aiNet and K-means (aiNet_KMC), a new one (aiNet_DP) based on density peaks (CDP) is presented to obtain a more accurate number of clusters and cluster centres according to the aiNet clustering results.

III. Proposal

The present work implements the artificial immune networks (aiNet) with two clustering algorithms, a one-pass clustering algorithm and another one based on density (see Figure 2). These two subsequent steps help redistribute the set of representatives generated by aiNet (a set of representatives for a single behaviour), thus making it able to detect various behaviours. A similar combination of technologies is already found in [27], where K-means is combined with DBSCAN and a genetic algorithm for image association.

In this approach, the goal of the aiNet algorithm is to model a generic behaviour, while the rest of the steps are devoted to redistributing the antibodies from that modelling. In the first step, from a set of data given as input, aiNet will return a set of representatives (detectors) that tries to model generically the behaviour of the data shown as input.

The problem is that this generic behavior models itself as one behaviour. To allow the identification of different subpatterns (that could be used for detecting anomalies), the redistribution of the representatives generated by aiNet is necessary. The main focus of this work is to get a generalized and refined version of aiNet, so the K-means algorithm is used as the second part of the process. For this, data is clustered by the algorithm, initializing it by passing the detectors generated by aiNet as initial centroids. In this sense, centroids (the detectors generated by aiNet) will be redistributed representing subpatterns by agrupating them.

Once the redistribution of the representatives has been made by K-means (1), a refinement process is necessary to group them by behaviour, thus allowing a correct identification of the represented subpatterns (2). For this purpose, the density-based clustering algorithm is used, which can be either DBSCAN or OPTICS. The set of detectors is passed as input and returns the same set together with the behavioural identifications of each detector.

However, in this case, density-based clustering methods are not used on the training data (as a usual clustering), not because it is inefficient (which can be), but because the proposed method looks for representatives of the sets (first global representatives with aiNet and then local with K-means). At this point, representatives of behaviours already exist, but it is not known which ones represent which. That is why DBSCAN is used to group the representatives by density to precisely represent different regions in the data.

After this, the behaviour given by DBSCAN or OPTICS is assigned to the detectors and the behaviour is equated to a label so that each behaviour will have an assigned label with which the model can be evaluated.

It should be noted that the aim is not to only cluster the detectors, but rather to determine the subpatterns (behaviours) that have been represented by them.

IV. Experimental Framework

In this section, a brief description of the dataset along with the data cleaning process is presented, after that some feature tuning recommendations are made and the performance metrics are introduced.

A. Dataset preprocessing and feature selection

The experiments drawn in this research are implemented using the *BoT-IoT* dataset[28], [29]. Created by designing a realistic network environment at UNSW Canberra's *Cyber Range Lab*, incorporates a combination of normal and *botnet* traffic. The extracted traffic stream, in CSV format, is 16.7 GB in size and the files are separated based on attack category and subcategory to better assist in the labelling process.

To facilitate the handling of the dataset, a 5% extracted from the original dataset (four files of approximately 3.6 million instances and 1.07 GB total size) was provided by Koroniotis et al. [28]. According to the authors, this subset is a representative sample of the full set in terms of attack category and has the most features of any set or subset processed from *Bot-IoT*, with 43 independent features and 3 dependent features. The 43 independent features contain Argus network flow features and additional computed features.

The final subset, named 10-*Best Subset*, contains the same number of instances as the original 5% *Subset* but contains only the 10 most important features. The top 10 features were derived by mapping the correlation coefficient and joint entropy of the 43 independent features and selected based on their ranking. As can be seen in Table I, the number of instances of each category is highly unbalanced, with DOS and DDOS attacks being predominant.

Table I
BoT-IoT: 5% Subset and 10-Best Subset

Category	Subcategory	Number of instances
Normal	Normal	477
DoS/DDoS	TCP	1,593,180
	UDP	1,981,230
	HTTP	2,474
Reconnaissance	OS Fingerprint	17,914
	Service Scanning	73,168
Information theft	Keylogging	73
	Data Exfiltration	6

Six invalid features and all `ICMP` and `ARP` values were removed following review and analysis by Peterson et al. [30]. The cited work points out all dependent, independent and invalid features that undermine the effectiveness of a predictive model because they contribute to *overfitting* and limit generalization. The features used for model training

are: *stddev, N_IN_Conn_P_SrcIP, N_IN_Conn_P_DstIP, min, mean, max, state_number, drate and srate*.

The labels used for the realization of the experiments are *category* and *subcategory*. As it can be seen in table I, the dataset used is highly unbalanced. To create a more balanced distribution, the predominant labels (*DoS, DDoS and Reconnaissance*) were separated into subgroups by combining *Category* and *Subcategory*. This new feature is hereafter referred to as *case*.

To generate the training and validation set, to know the robustness of the approach, a Stratified K Fold Split (SKFS) was performed to ensure that the approach used was functional given the different distributions of the dataset. Once all the experiments were completed, the results obtained were averaged to represent the reliability and robustness of the experiments.

B. Density-based clustering models parameters

While in this work, K-means has an automatic hyperparameter setting, for aiNet, OPTICS and DBSCAN must be set manually. To select the optimal values for focus validation with the presented dataset, several guidelines have been followed. For aiNet, the predetermined hyperparameter values have been used. The division of subpatterns made by DBSCAN and OPTICS is adapted to each case. Given that it is known that the dataset has 10 different labels, it is believed that the optimal hyperparameters for the validation of the model will be the one that presents 10 subpatterns, reflecting the labels of the dataset in Table I.

However, for the behaviour identification problem, an optimal number does not exist, since a greater or lesser number of identified subpatterns simply indicates a greater or lesser criteria). Models such as DBCSCAN or OPTICS can identify points that do not belong to a cluster, these are assigned the value -1. However, in this case, not the principal data set but the set of detectors generated by aiNet and relocated by K-means is introduced as input. In this case, assuming that all the detectors can provide information in their location, a process of recovery of the detectors with value -1 was chosen, reassigning the identified behaviour to preserve the original number of detectors and therefore, the highest possible results.

This process of resignation involves two proposals: The first and simplest is the assignment of the behaviour of the nearest neighbouring detector (performing a calculation of the minimum distance of that detector to the others. The second is to assign the behaviour to the cluster with the smallest average distance to its points. In other words, after calculating the average distance from the detector with the -1 value to the other detectors in the same cluster, the behaviour of the closest cluster is assigned.

For validation purposes and to ensure the representatives generated by our model identify the subpatterns that exist in the training data, a two-step assignation is proposed to link a known label (the ground truth) to a subpattern/behaviour (identified by DBSCAN or OPTICS). First, the average distance from an individual detector to all the points of the same ground truth cluster is calculated. Then, the smallest average distance is assigned to that detector. This procedure alone does not make use of the information obtained by density-based clustering methods, since it treats detectors as independent individuals and not as points of a grouped behaviour. Therefore, to take advantage of such information, this approach carries an added step of *majority label reassignment*, in which, after completing the process described above, the predominant label is assigned for all detectors of the same behaviour (calculated using DBSCAN/OPTICS).

C. Evaluation Metrics

Clustering is a widely used unsupervised process that is especially sensitive to the input parameters and therefore it is important to evaluate the results of clustering algorithms. However, it is difficult to define when a clustering result is acceptable. In this work, external validation techniques (using the available label of the data) are performed.

After the proposed algorithms group the data, it is possible to differentiate several concepts: *True Positive Value* (*VP*) refers to those points that were placed by the algorithm in the same cluster that indicated the class that was counted beforehand. *False Positive* (*FP*) refers to those points that were placed by the algorithm in a cluster and that belonged to another cluster. *False Negatives* or (*FN*) refers to those elements of a cluster that were placed in a different cluster than the one indicated by its label. *True Negative* (*VN*) refers to those elements that were correctly placed outside a cluster.

$$Precision = \frac{VP}{VP + FP} \ , \tag{1}$$

$$Recall = \frac{VP}{VP + FN} \ , \tag{2}$$

$$F_\alpha = \frac{1+\alpha}{\frac{1}{precision} + \frac{\alpha}{recall}} \ . \tag{3}$$

With these values, it is possible to introduce the following widely used metrics: *precision Ec. (1)* (measures the relative success rate of the model referred to the total amount of real positives), *recall Ec. (2)*(measures the relative success rate of the model referred to the the total amount of predicted positives), *F1-score Ec. (3)* (the harmonic mean of both) and support (the number of instances with that class used in the validation set).

V. Experimental Results

Table II
Experimental results with AINET+KMC+DBSCAN

Label	Precision	Recall	F1-Score	Support
0	1	1	1	28150
1	1	1	1	16794
2	1	0.97	0.99	37
3	0.9	1	0.95	25667
4	1	0.83	0.9	26826
5	0.01	1	0.03	24
6	1	1	1	1769
7	1	1	1	474
8	0.85	1	0.92	11
9	0	0	0	2
Weighted avg	0.97	0.95	0.96	99754

The configurations formed by aiNet_KMC together with DBSCAN offer an average accuracy of 0.95 (see Table II. In the case of the configurations formed by aiNet_KMC together

Table III
EXPERIMENTAL RESULTS WITH AINET+KMC+OPTICS

Label	Precision	Recall	F1-Score	Support
0	1	1	1	28150
1	1	1	1	16794
2	0	0	0	37
3	0.9	0.96	0.93	25667
4	0.96	0.83	0.89	26826
5	0.01	0.88	0.02	24
6	0.98	0.26	0.41	1769
7	0.29	0.98	0.45	474
8	0.85	1	0.92	11
9	0	0	0	2
Weighted avg	0.96	0.93	0.94	99754

with OPTICS, slightly lower results are obtained with an accuracy of 0.93 (see Table III).

In both cases, while most behaviours are detected correctly, labels with lower instances are not. This may be because either not enough amount of instances have been used in the training phase (thus not pulling any detectors to locations for those identifications to happen correctly) or there exist similarities between clusters (i.e. DDOS/DOS via TCP or UDP).

Even so, the presented model can still be considered valid. After analyzing the results it is concluded that the identification of behavioral representatives is performed correctly. Considering these results, it is observed that the use of DBSCAN has achieved higher performance than other similar unsupervised clustering methods like UADINK (aiNet and K-Means combination) [13] and ADAID (aiNet and Density Peaks combination) [26] in terms of the correct classification of instances in the validation and training test. It is necessary to highlight that although both DBSCAN and OPTICS have had a good performance, the adjustment of their hyperparameters has been performed manually, so it would be interesting to consider the automation of this process.

VI. CONCLUSIONS AND FUTURE WORK

The proposed work allows obtaining detectors in different regions from clustering and classifying behaviours in the context of cybersecurity. This model is understood as an extension of existing technologies such as artificial immune networks. The combination of aiNet with K-means has allowed refining the identification of sub-patterns, while the use of density-based clustering techniques such as DBSCAN and OPTICS has allowed the construction of behaviours from the sub-patterns.

Although labels with smaller instances are not labelled correctly, the experimental result on the described dataset confirms that the proposed three-stage approach can help evaluate randomly shaped data similar to UADINK [13] and ADAID [26].

There are several steps in the work to be continued and improved, such as the comparison of the actual method with other state-of-the-art options and performing an ablation study without the K-means step.

Another improvement would be the inclusion of negative detectors (not anomaly but normality detectors) in aiNet implementation to improve the detection and classification of possible attacks by generating a secure profile. By proposing the modification of aiNet, a crossover operator between detectors could be introduced that would cause the generation of new descendants that can correctly interpret those points that are misidentified with the current population. Other contributions are the selection of metrics for internal validation without the need for a labelled dataset.

ACKNOWLEDGEMENTS.

The work presented is carried out within the Digital Security area of Vicomtech within the framework of the CERVERA network of national excellence for research in information privacy technologies, ÉGIDA (EXP 00122721 / CER-20191012) - RED DE EXCELENCIA EN TECNOLOGIAS DE SEGURIDAD Y PRIVACIDAD and the regional projects of applied basic research REMEDY (KK-2021/00091) and BEACON (KK-2023/00085).

REFERENCES

[1] D. Evans, "The internet of things: How the next evolution of the internet is changing everything," *CISCO White Paper*, vol. 1, pp. 1–11, 2011.
[2] S. Ray, Y. Jin, and A. Raychowdhury, "The changing computing paradigm with internet of things: A tutorial introduction," *IEEE Design and Test*, vol. 33, pp. 76–96, 2016.
[3] E. Bertino and N. Islam, "Botnets and internet of things security," *Computer*, vol. 50, no. 2, pp. 76–79, 2017.
[4] S. Raza, L. Wallgren, and T. Voigt, "Svelte: Real-time intrusion detection in the internet of things," *Ad Hoc Networks*, vol. 11, no. 8, pp. 2661–2674, 2013.
[5] P. Dixit, R. Kohli, A. Acevedo-Duque, R. R. Gonzalez-Diaz, and R. H. Jhaveri, "Comparing and analyzing applications of intelligent techniques in cyberattack detection," *Security and Communication Networks*, vol. 2021, 2021.
[6] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," *Applied Sciences*, vol. 9, p. 4396, 2019.
[7] J. Meakins, "A zero-sum game: the zero-day market in 2018," *Journal of Cyber Policy*, vol. 4, no. 1, pp. 60–71, 2019.
[8] P. Casas, J. Mazel, and P. Owezarski, "Unsupervised network intrusion detection systems: Detecting the unknown without knowledge," *Computer Communications*, vol. 35, no. 7, pp. 772–783, 2012.
[9] K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," *Proceedings of the Twenty-Eighth Australasian Conference on Computer Science*, vol. 38, pp. 333–342, 2005.
[10] X. Niu, L. Wang, and X. Yang, "A comparison study of credit card fraud detection: Supervised versus unsupervised," 2019.
[11] L. N. de Castro and F. J. V. Zuben, "ainet: An artificial immune network for data analysis," *IGI Global*, pp. 231–260, 2001.
[12] N. Jerne, "Towards a network theory of the immune system," *Annales d'Immunologie*, vol. 125C, no. 1-2, pp. 373–389, 1974.
[13] Y. Shi, X. Peng, R. Li, and Y. Zhang, "Unsupervised anomaly detection for network flow using immune network based k-means clustering," *Communications in Computer and Information Science*, pp. 386–399, 2017.
[14] S. P. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, 1982.
[15] J. MacQueen, "Classification and analysis of multivariate observations," *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.
[16] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, why and how you should (still) use dbscan," *ACM Transactions on Database Systems*, vol. 42, no. 3, pp. 1–21, 2017.
[17] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "Optics: Ordering points to identify the clustering structure," *Association for Computing Machinery*, pp. 49–60, 1999.
[18] G. Münz, S. Li, and G. Carle, "Traffic anomaly detection using k-means clustering," *GI/ITG Workshop MMBnet*, vol. 7, p. 9, 2007.
[19] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 303–336, 2014.
[20] P. Laskov, P. Düssel, C. Schäfer, and K. Rieck, "Learning intrusion detection: Supervised or unsupervised?" *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3617, pp. 50–57, 2005.

[21] Y. Li, D. Wang, Y. Yu, and L. Jiao, "An improved artificial immune network algorithm for data clustering based on secondary competition selection," *2016 IEEE Congress on Evolutionary Computation, CEC 2016*, pp. 2744–2751, 11 2016.

[22] T. Stibor and J. Timmis, "An investigation on the compression quality of ainet," *Proceedings of the 2007 IEEE Symposium on Foundations of Computational Intelligence, FOCI 2007*, pp. 495–502, 2007.

[23] H. Lau, J. Timmis, and I. Bate, "Anomaly detection inspired by immune network theory: A proposal," *IEEE congress on evolutionary computation*, pp. 3045 – 3051, 2009.

[24] M. Rassam and M. Maarof, "Artificial immune network clustering approach for anomaly intrusion detection," *Journal of Advances in Information Technology*, vol. 3, pp. 147–154, 2012.

[25] Y. Shi, R. Li, Y. Zhang, and X. Peng, "An immunity-based time series prediction approach and its application for network security situation," *Intelligent Service Robotics*, vol. 8, pp. 1–22, 2015.

[26] Y. Shi and H. Shen, "Anomaly detection for network flow using immune network and density peak," *International Journal of Network Security*, vol. 22, pp. 337–346, 2020.

[27] S. Kurumalla and C. Kanda, "Dbscan assisted by hybrid genetic k means algorithm," *International Journal of Recent Technology and Engineering*, vol. 8, pp. 1973–1979, 2020.

[28] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset," *Future Generation Computer Systems*, vol. 100, pp. 779–796, 2018.

[29] N. Koroniotis, N. Moustafa, E. Sitnikova, and J. Slay, "Towards developing network forensic mechanism for botnet activities in the iot based on machine learning techniques," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, vol. 235, pp. 30–44, 2018.

[30] J. M. Peterson, J. L. Leevy, and T. M. Khoshgoftaar, "A review and analysis of the bot-iot dataset," *Proceedings - 15th IEEE International Conference on Service-Oriented System Engineering, SOSE 2021*, pp. 20–27, 2021.