IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Darknet Traffic Detection System Using Feature Selection Techniques, Analysis of Machine Learning Classifiers

Bhavesh Patil
Department of Computer Engineering,
RMD Sinhgad School of Engineering,
Pune, India

Prof. Jyoti Raghatwan
Department of Computer Engineering,
RMD Sinhgad School of Engineering,
Pune, India

Saurabh Sarate
Department of Computer Engineering,
RMD Sinhgad School of Engineering,
Pune, India

Bushra Tamboli
Department of Computer Engineering,
RMD Sinhgad School of Engineering,
Pune, India

Abstract— The Darknet, a portion of the deep web, has seen an increase in illegal activities such as drug trafficking, terrorism, extremism, and child pornography. Therefore, there is a pressing need for proper identification and classification of Darknet traffic. The purpose of this paper is to demonstrate that the random forest classifier is the most suitable algorithm for classifying dark- net traffic. This will be achieved by comparing its performance metrics with several other machine learning algorithms such as an Extra tree, AdaBoost, SVM, decision tree, etc., which were found to have the highest accuracy of 98.14%.

Index Terms—Darknet, machine learning techniques, classification

I. INTRODUCTION

Darknet is a term used to refer to a portion of the internet that is not indexed by search engines and requires specific software and configurations to access. It is often used for purposes of anonymity, privacy, and security.[13] However, in recent years, the Darknet has also become known for its association with illegal activities such as drug trafficking, terrorism, extremism, and child pornography. The anonymity and lack of oversight on the Darknet have made it a safe haven for criminals to carry out their activities. As a result, there is a growing need for proper identification and classification of Darknet traffic to combat these illegal activities. In this paper, we present the successful application of machine learning techniques to distinguish Darknet traffic from safer ones and identify the type of application running beneath the Darknet traffic. Our results offer valuable insights into identifying and classifying Darknet traffic, which can aid in the fight against illegal activities on the Darknet.[5]

TOR (The Onion Router) is open-source software that enables users to browse the internet anonymously by channeling their traffic through servers run by volunteers. The TOR network is frequently utilized to reach the darknet, which encompasses networks and websites that aren't included in the conventional internet and aren't indexed by search engines 1 [3]. Non-TOR networks in the darknet allude to networks and websites that exist beyond the conventional internet and can't be accessed through a TOR network, but alternative methods like peer-to-peer networks, customized protocols, or other anonymity services can be employed. VPN (Virtual Private Network) is a technology that facilitates secure and encrypted connections for a user to a remote server, through which their internet traffic is directed. This provides a user with the ability to conceal their location and IP address while encrypting their traffic. VPNs are often utilized to safeguard privacy and security while using the internet and can also be used to access the darknet. Non-VPN networks in the darknet are networks and websites that exist outside of the conventional internet and can't be accessed through VPNs but can be accessed through other methods like the TOR network or other anonymity services.

The traffic in the darknet comprises a blend of lawful and illicit traffic, encompassing traffic from the Tor network, which is frequently used to access the darknet. Among the frequently used types of network traffic are TOR, Non-TOR, VPN, and Non-

VPN.[2]

II. RELATEDWORK

Research Gaps

1. In consistent dataset: While there are many publicly available datasets for network traffic analysis, there is a lack of a standardized dataset for Darknet traffic classification. This makes it difficult to compare the performance of different feature selection and classification techniques.
2. Least research on the impact of different feature selection techniques: While many different feature selection techniques have been proposed for network traffic classification, there is limited research on the impact of using different feature selection techniques on classification accuracy. More research is needed to understand which feature selection techniques work best for Darknet traffic classification.
3. Lack of information related to impact of different classification algorithms: While various machine learning algorithms have been proposed for network traffic classification, there is limited research on the impact of using different classification algorithms for Darknet traffic classification. More research is needed to understand which classification algorithms work best for Darknet traffic classification.
4. Lack of research on the generalizability of the models: Most research on Darknet traffic classification using feature selection has focused on building models for a specific Darknet dataset. However, it is important to determine the generalizability of the models to other Darknet datasets and real-world network traffic.
5. Limited research on the impact of feature selection on model interpretability: While feature selection can improve classification accuracy, it can also make the resulting models less interpretable. There is a need for research to understand the trade-off between model interpretability and classification accuracy when using feature selection for Darknet traffic classification.

III. LITURATURE SURVEY

1. “A Machine Learning Approach to Classify Network Traffic”

This paper analyzes CIC-Darknet 2020 dataset to classify the benign and darknet traffic. Before applying any classifiers to our dataset, we have balanced it using Synthetic Minority Oversampling Technique (SMOTE). We have applied PCA to reduce dimensionality, furthermore, ensemble techniques, logistic classifiers, tree-classifiers, and Naive Bayes have been compared and evaluated thoroughly with various evaluation metrics Accuracy, Precision, Recall, F1-Score, and Mathew’s Correlation Coefficient (MCC).

2. “Darknet Traffic Classification using Machine Learning Techniques”

A Darknet is an overlay network within the Internet, and packets' traffic originating from it is usually termed as suspicious. In this paper common machine learning classification algorithms are employed to identify Darknet traffic. A ROC analysis along with a feature importance analysis for the best classifier was performed, to provide a better visualization of the results.

3. “Darknet Traffic Classification with Machine Learning Algorithms and SMOTE Method”

In this paper, we proposed three different machine learning (ML) based traffic classification approaches; the binary classification of Darknet and Benign traffic classes; the quadruple classification of classes Tor, Non-Tor, VPN, and Non-Vpn; an traffic classification of eight sub-traffic classes. We further applied the SMOTE method for balancing the sizes of the classes in the traffic dataset and feature selection (FS) algorithms to identify the most effective attributes where the number of features in the original dataset were reduced from 63 to 8, 8 and 6.

4. “DIDarknet: A Contemporary Approach to Detect and Characterize the Darknet Traffic using Deep Image Learning”

Darknet traffic classification is significantly important to categorize real-time applications. Although there are notable efforts to classify darknet traffic which rely heavily on existing datasets and machine learning classifiers, there are extremely few efforts to detect and characterize darknet traffic using deep learning.

5. “DarkNet Traffic Classification Pipeline with Feature Selection and Conditional GAN-based Class Balancing”

In this paper, the standard CIC-Darknet2020 dataset used contained instances of benign and DarkNet traffic to a network. Feature importance analysis is performed using Chi-Squared statistical score on the dataset to aid in feature selection. The imbalance of the classes is then handled by performing oversampling using Conditional Generative Adversarial Networks. The multi-class classification of the traffic encryption type is performed using Random Forest classifier. This pipeline performs with a F1-Score of 97.87 for traffic encryption classification

6. “Towards Early Detection of Novel Attack Patterns through the Lens of a Large Scale Darknet”

Darknet monitoring provides a cost-effective way to monitor the global trend of cyber-threats in the Internet. To make full use of the darknet traffic at hand, in this paper, we present a study on early detection of emerging novel attacks observed in the darknet. First, exploration of the regularities in the communications from attacking hosts are done by feeding all observed packets in the darknet to a frequent itemset mining engine, where the most frequently occurred attack patterns are automatically grouped together. Second, a time series which characterizes the activity level of each attack pattern is created over the observation period. Then, to extract the most prominent attack patterns, a clustering algorithm is engaged to cluster the attack patterns into groups that carry the similar activities in a long run, dimension reduction is employed to provide visual hints about their relationship.

7. “Machine-Learning-Based Darknet Traffic Detection System for IoT Applications”

The massive modern technical revolution in electronics, cognitive computing, and sensing has provided critical infrastructure for the development of today’s Internet of Things (IoT) for a wide range of applications. However, because endpoint devices’ computing, storage, and communication capabilities are limited, IoT infrastructures are exposed to a wide range of cyber-attacks. As such, Darknet or blackholes (sinkholes) attacks are significant, and recent attack vectors that are launched against several IoT communication services. Since Darknet address space evolved as a reserved internet address space that is not contemplated to be used by legitimate hosts globally, any communication traffic is speculated to be unsolicited and distinctively deemed a probe, backscatter, or misconfiguration. Thus, in this paper, we develop, investigate, and evaluate the performance of machine-learning based Darknet traffic detection systems (DTDS) in IoT networks.

8. “Robust stacking ensemble model for darknet traffic classification under adversarial settings”

Encrypted traffic tunneled by Tor or VPN is referred to as darknet traffic. The ability to detect, identify, and characterize darknet traffic is critical for detecting network traffic generated by a cyber-attack. Darknet classification models based on Machine Learning / Deep Learning (ML/DL) usually demonstrate high False Positive Rate (FPR) and lower F1-score which are essential metrics for network traffic analysis. Additionally, ML/DL models used in such tasks are susceptible to adversarial perturbed samples that can cause the network security solution to malfunction. This work proposes a Stacking Ensemble (SE) model to combine the predictions of three base learners, 1) Random Forest, 2) K-Nearest Neighbors, and 3) Decision Tree in the most efficient way for improving the overall performance of darknet characterization.

9. “Darknet traffic big-data analysis and network management for real-time automating of the malicious intent detection process by a weight agnostic neural networks framework.”

Attackers are perpetually modifying their tactics to avoid detection and frequently leverage legitimate credentials with trusted tools already deployed in a network environment, making it difficult for organizations to proactively identify critical security risks. Network traffic analysis products have emerged in response to attackers’ relentless innovation, offering organizations a realistic path forward for combatting creative attackers. Additionally, thanks to the widespread adoption of cloud computing, Device Operators (DevOps) processes, and the Internet of Things (IoT), maintaining effective network visibility has become a highly complex and overwhelming process. What makes network traffic analysis technology particularly meaningful is its ability to combine its core capabilities to deliver malicious intent detection. In this paper, we propose a novel darknet traffic analysis and network management framework to real-time automating the malicious intent detection process, using a weight agnostic neural networks architecture.

10. “Classification of VPN network traffic flow using time related features on Apache Spark”

This paper classifies the VPN network traffic flow using the time related features on the Apache Spark and artificial neural networks. Today's, internet traffic is encrypted using protocols like VPN/Non-VPN. This situation prevents the classic deep packet inspection approaches by analyzing packet payloads. For the implementation of this research, MATLAB 2019b would be forwarded in use as increasing demand for VPN networks has actuated the evolutionary technology. The proposed method will prevent unnecessary processing as well as flooding found in standard VPN network traffic classification. As the proposed system is trained on 80 of the datasets while 20% is kept for the testing and validation with 10-cross fold validation as well as 50 epochs of training. To the best of our knowledge, this is the first study that introduces and utilizes artificial neural networks and apache spark engine to implement the classification of VPN network traffic flow.

The accuracy of the VPN classification using ANN and Apache Spark Engine is 96.76%. The accuracy of the Non-VPN classification using the proposed method is 92.56%. This study has shown that an approach using the CICDarknet2020 for packet-level encrypted traffic classification cannot incorporate packet header information, as it allows to directly map a packet to a specific application with high accuracy.

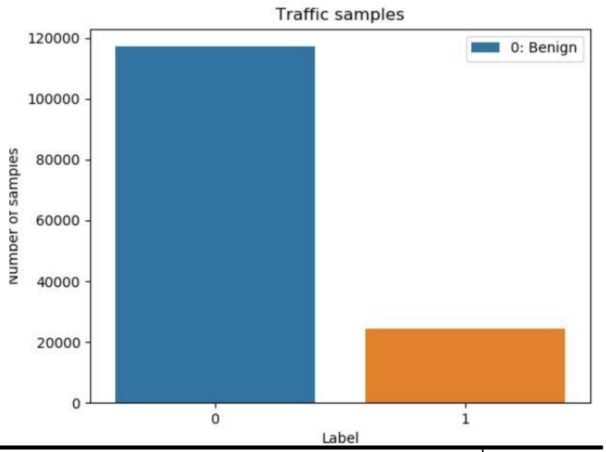
11. “Multimodality data analysis in information security ETCC: encrypted two-label classification using CNN”

Due to the increasing variety of encryption protocols and services in the network, the characteristics of the application are very different under different protocols. However, there are very few existing studies on encrypted application classification considering the type of encryption protocols. In order to achieve the refined classification of encrypted applications, this paper proposes an Encrypted Two-Label Classification using CNN (ETCC) method, which can identify both the protocols and the applications. ETCC is a two-stage two-label classification method

IV. PROPOSED METHODOLOGY

A. Dataset

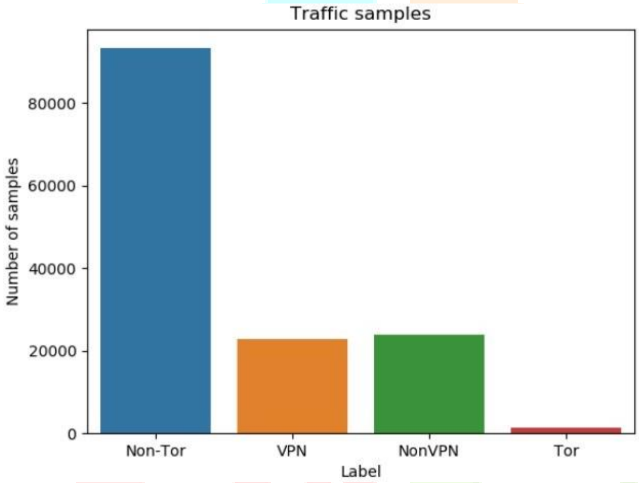
In this study, the CIC-Darknet2020 dataset[2][1] was utilized, which contains 141530 rows and 85 columns of features.



Some unnecessary features were eliminated during prepro- cessing. The label feature with four labels, TOR, VPN, non- TOR, and non-VPN, was used for classification. Benign traffic includes TOR and VPN, while darknet traffic comprises non- TOR and non-VPN, indicating a higher risk of attack. Fig 1 displays that the dataset includes 117170 benign samples and 24311 darknet samples. Further analysis of the labels revealed 93356 samples of non-TOR, 23863 samples of non-VPN, 22917 samples of VPN, and 1392 samples of TOR, as shown in Fig 2.

B. Data Pre-processing

The dataset used in this study comprises 141530 rows, but some of these rows contain NaN or infinite values, which were handled in the preprocessing step. After this step, the dataset was reduced to 141480 rows, and 79 features were used for traffic classification. The NaN values were replaced with 0 during preprocessing.



C. Machine Learning Algorithms

Machine learning (ML) is a computer science field that uses data to learn and generate predictions or inferences. The process is possible using mathematical components like statistics, probability distribution, and differential calculus.[11] ML can be divided into three distinct categories: supervised, unsupervised, and reinforcement learning. However, the cat- egorization heavily relies on the type of input needed to generate the intended output. The efficacy of the machine learning approach is largely dependent on the quality of the dataset used.

Supervised learning involves working with a labeled dataset, which means that the data has feature labels. The objective of this approach is to identify and comprehend patterns within the data, in order to generate an output.[12]These algorithms are designed to recognize complex patterns in large datasets and then use those patterns to make predictions or decisions. There are many machine learning algorithms that can be used with feature selection techniques to improve their performance. Some popular algorithms include Decision Trees, Random Forest, Support Vector Machines, and others. These algorithms can use feature importance measures, such as information gain or mutual information, to identify the most relevant features in the dataset.

D. Evaluation Metrics

We utilized several evaluation metrics in our study, including accuracy, precision, recall, and F1-score.[1]

1) *Accuracy:* Accuracy is an evaluation metric used to measure the proportion of correct predictions made by a model, typically expressed as a percentage. It is calculated by dividing the number of correct predictions by the total number of predictions made.

$$Accuracy= \frac{TP + TN}{TP + TN + FP + FN}$$

2) *Recall:* Recall is an evaluation metric used to measure the proportion of positive instances that are correctly identified by a model. It is calculated by dividing the true positive predictions by the total number of actual positive instances

3) *Precision:* Precision is an evaluation metric used to measure the proportion of positive predictions that are correct. It is calculated by dividing the true positive predictions by the total number of positive predictions made

$$Precision = \frac{TP}{TP + FP}$$

4) F- score: F-score is an evaluation metric used to measure the balance between precision and recall in a model's predictions. It is the harmonic mean of precision and recall and ranges from 0 to 1 , where 1 indicates the best possible performance.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

5) HyperParameters Setting

- Training size = 80 %, Testing size = 20%
- For Random Forest: Max depth = 19

TABLE I
DATASET SPLITING

Train size(%)	Test size(%)	Accuracy
10	90	96.896614
20	80	97.448421
30	70	97.737986
40	60	97.857934
50	50	97.966509
60	40	98.062248
70	30	98.040463
80	20	98.14527
90	10	98.106409

In our experiment, we were interested in evaluating the perfor- mance of a random forest model on a dataset. To do this, we first divided the dataset into two sets: training and testing, with a different ratio for each. Specifically, we allocated 10% of the data to the testing set and the remaining 90% to the training set. We then trained the random forest model on the training set, with its max depth attribute set to 19, and evaluated its performance on the testing set in terms of accuracy. We repeated this process for each ratio and found that the performance of the model varied depending on the ratio of training and testing data. Interestingly, we found that the highest accuracy (98.14%) was achieved when using an 80-20% split for training and testing. This means that when we trained the model on 80% of the data and tested it on the remaining 20%, it performed the best in terms of accuracy.

As a result of this finding, we decided to use a fixed ratio of 80% for training and 20% for testing in all subsequent experiments. This ensures that we are using the ratio that gave us the best performance in our initial experiment, allowing for more accurate comparisons between different models or experimental conditions in future studies.

V. Classifiers Output

TABLE II
Summary of Random Forest

Train(%)	Test(%)	Accuracy	Precision	Recall	F1-score	Time(s)
90	10	98.1064	98.1139	98.1064	98.1051	24.6577
80	20	98.1452	98.1459	98.14527	98.1426	21.7836
70	30	98.0404	98.0411	98.0404	98.0376	19.8898
60	40	98.0622	98.0643	98.0622	98.0599	16.2550
50	50	97.9650	97.9682	97.9650	97.9633	13.9904
40	60	97.8579	97.8579	97.8579	97.8555	11.1743
30	70	97.7369	97.7369	97.7396	97.7338	9.0384
20	80	97.4484	97.4496	97.4484	97.4437	6.2778
10	90	96.8973	96.8953	96.8973	96.8930	4.3387

TABLE III

Summary of Decision Tree Classifier

Train(%)	Test(%)	Accuracy	Precision	Recall	F1-score	Time(s)
90	10	97.79552	97.7998	97.7955	97.7951	2.6264
80	20	97.73193	97.7322	97.7319	97.7307	2.2984
70	30	97.6400	97.6436	97.6400	97.6403	2.0522
60	40	97.4969	97.5005	97.4969	97.4978	1.6711
50	50	97.4252	97.4372	97.4252	97.4294	1.4208
40	60	97.3503	97.3534	97.3503	97.3512	1.1308
30	70	96.8880	96.8832	96.8880	96.8851	1.5428
20	80	96.7498	96.7537	96.7498	96.7517	0.6067
10	90	96.3148	96.3290	96.3148	96.3166	0.3789

TABLE IV
Summary of Extra Tree Classifier

Train(%)	Test(%)	Accuracy	Precision	Recall	F1-score	Time(s)
90	10	97.7531	97.8852	97.7531	97.8183	27.4043
80	20	97.7531	97.9443	97.7531	97.8479	24.5780
70	30	97.6989	97.8833	97.6989	97.7902	22.0064
60	40	97.6489	97.8643	97.6489	97.7554	17.2683
50	50	97.5892	97.8191	97.5892	97.7033	14.9285
40	60	97.4186	97.6784	97.4186	97.5471	12.6080
30	70	97.2736	97.6062	97.2736	97.4378	10.7303
20	80	97.0386	97.4228	97.0386	97.2275	8.1420
10	90	96.3054	96.8987	96.3054	96.5975	6.6008

VI. RESULTS AND DISCUSSION

The experiment involved applying several machine learning models, including random forest, KNN, SVC, Gradient Boost, Extra Tree, and Decision Tree, to the dataset. The results showed that the random forest model, with a maximum depth of 19, performed the best among all the models with an accuracy of 98.14527%. On the other hand, the decision tree model came in second place, with an accuracy of 97.79552%. However, it took slightly longer to process, increasing from approximately 21 to 26 seconds, as shown in Table.

TABLE V

MODEL PERFORMANCE

Classifier	Train size	Test size	Accuracy	Precision	Recall	F1-score	Elapsed time to compute
Random Forest	80	20	98.1452	98.1459	98.1452	98.1426	21.7836
Decision Tree	90	10	97.7955	97.7998	97.7952	97.7951	26.2643
Extra Tree	80	20	97.7531	97.9443	97.7531	97.8479	24.5780
KNN	90	10	92.6799	92.9160	92.6799	92.6847	32.8560
SVC	80	20	66.1273	57.0594	66.1273	53.1791	1977.1385
Gradient Boost	50	50	96.6876	96.6823	96.6876	96.6821	183.6799

TABLE VI
Summary of KNN

Train(%)	Test(%)	Accuracy	Precision	Recall	F1-score	Time(s)
90	10	92.6799	92.9160	92.6799	92.6847	32.8560
80	20	92.4256	92.5383	92.4256	92.3822	55.1448
70	30	92.1524	92.2276	92.1524	92.1105	72.3029
60	40	92.3161	92.4341	92.3161	92.2702	85.6634
50	50	92.0426	92.1387	92.0426	92.0202	91.4836
40	60	91.8486	91.9519	91.8486	91.8160	83.1786
30	70	91.2335	92.2775	91.2335	91.6198	81.6639
20	80	91.2792	91.4298	91.2792	91.1978	61.0984
10	90	90.7212	91.1774	90.7212	90.6901	38.3706

The SVC model showed the least performance among all the models, indicating that it may not be the best choice for this particular task. In comparison, the Extra Tree and Decision Tree models had a small difference in their accuracy, but the Extra Tree model required less time to process and used a smaller training dataset size.

Score Function	No of Features	Time in(s)	Accuracy	Precision	Recall	F1-Score
No-Score	79	22.1142931	98.14527	98.145988	98.14527	98.142617
GINI	65	24.52691102	98.297181	98.300598	98.297181	98.294588
ANOVA F-test	45	20.88681364	98.159401	98.163321	98.159401	98.156945
Mutual Info	30	20.09399939	98.152335	98.153207	98.152335	98.149217

TABLE VII
SUMMARY OF SVC

Train(%)	Test(%)	Accuracy	Precision	Recall	F1-score	Time(s)
90	10	66.0425	56.7525	66.0425	53.1268	1488.2913
80	20	66.1273	57.0594	66.1273	53.1791	1977.1385
70	30	65.4254	42.8049	65.4254	51.7513	1851.2593
60	40	65.5567	42.9769	65.5567	51.9180	939.7998
50	50	65.6694	43.1247	65.6694	52.0612	785.6123
40	60	65.7822	43.2730	65.7822	52.2047	599.5651
30	70	65.8053	43.3034	65.8053	52.2340	453.972603
20	80	65.8941	43.4203	65.8941	52.3471	303.9517
10	90	65.9145	43.4473	65.9145	52.3731	134.8091

In summary, the results suggest that the random forest model with a maximum depth of 19 is the most effective for this particular task, with the decision tree model being a close second. However, the Extra Tree model may be a more efficient option in terms of processing time and training dataset size.

TABLE VIII
SUMMARY OF GRADIENT BOOSTING CLASSIFIER

Train(%)	Test(%)	Accuracy	Precision	Recall	F1-score	Time(s)
90	10	96.6155	96.6202	96.6155	96.6140	328.0303
80	20	96.6014	96.6054	96.6014	96.6000	293.7699
70	30	96.4860	96.4905	96.4860	96.4841	253.5546
60	40	96.6650	96.6669	96.6650	96.6627	216.8384
50	50	96.6876	96.6823	96.6876	96.6821	183.6799
40	60	96.6096	96.5960	96.6096	96.6010	142.8969
30	70	96.4727	96.4662	96.4722	96.4672	104.8402
20	80	96.4813	96.4707	96.4813	96.4743	67.41712
10	90	96.3212	96.3073	96.3219	96.3140	33.31418

Ranking Approach

We utilized three scoring functions, namely the Gini index, ANOVA F-test, and Mutual Info, as part of the Ranking Approach for feature selection. Our analysis showed that using the Gini index resulted in a reduction of 65 features, while the ANOVA F-test and Mutual Info resulted in reductions of45 and 30 features, respectively.

Table: Summary of Evaluation Metrics based on Ranking

Upon further analysis, we found that the Mutual Info technique had the minimum number of reduced features and also required the least amount of time. Therefore, we consider the Mutual Info scoring function to be the best approach for feature selection in classification tasks.

VII. CONCLUSION

In conclusion, the Darknet has become a hub for illegal activities, making it essential to properly identify and classify its traffic. This paper aims to identify the most suitable machine learning algorithm for this task and found that the random forest classifier outperformed other algorithms such as extra-tree, ada-boost, SVM,

and decision tree, achieving a maximum accuracy of 98.14%. Therefore, the random forest classifier is recommended as an effective method for the classification of Darknet traffic. This study can have significant implications in the field of cybersecurity, providing a reliable approach for identifying and preventing illegal activities on the Darknet.

VIII. FUTURE SCOPE

In order to gain a more comprehensive understanding, it is necessary to conduct a more advanced examination of the traffic characteristics within the Darknet as a part of our future work, we aim to enhance accuracy by employing various feature selection methods and efficient machine learning algorithms to reduce the number of features involved.

IX. REFERENCE

[1] Jadav, Nilesh, et al. ”A machine learning approach to classify network traffic.” 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE, 2021.

[2] Iliadis, Lazaros Alexios, and Theodoros Kaifas. ”Darknet traffic classification using machine learning techniques.” 2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST). IEEE, 2021.

[3] Karag¨ol, Hasan, et al. ”Darknet Traffic Classification with Machine Learning Algorithms and SMOTE Method.” 2022 7th International Conference on Computer Science and Engineering (UBMK). IEEE, 2022.

[4] Habibi Lashkari, Arash, Gurdip Kaur, and Abir Rahali. ”Didarknet: A contemporary approach to detect and characterize the darknet traffic using deep image learning.” 2020 the 10th International Conference on Communication and Network Security. 2020.

[5] Sridhar, Sashank, and Sowmya Sanagavarapu. ”DarkNet Traffic Classification Pipeline with Feature Selection and Conditional GAN-based Class Balancing.” 2021 IEEE 20th International Symposium on Network Computing and Applications (NCA). IEEE, 2021. 25

[6] Mohanty, Hardhik, Arousha Haghighian Roudsari, and Arash Habibi Lashkari. ”Robust stacking ensemble model for darknet traffic classification under adversarial settings.” Computers Security 120 (2022): 102830.

[7] Ul Alam, M. Z., A. Azizul Hakim, and M. Toufikuzzaman. ”Application and Interpretation of Ensemble Methods for Darknet Traffic Classification.” Preprint. In Proceedings of the 42nd IEEE Symposium on Security and Privacy, San Francisco, CA, USA. 2021.

[8] Demertzis, Konstantinos, et al. ”Darknet traffic big-data analysis and network management for real-time automating of the malicious intent detection process by a weight agnostic neural networks framework.” Electronics 10.7 (2021): 781.

[9] Aswad, Salma Abdullah, and Emrullah Sonu¸c. ”Classification of VPN network traffic flow using time related features on Apache Spark.” 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2020.

[10] Li, Yan, and Yifei Lu. ”Multimodality data analysis in information security ETCC: encrypted two-label classification using CNN.” Security and Communication Networks 2021 (2021): 1-11.