Darknet Traffic Detection based on Contrastive Learning

Daeun Kim, Yiseul Choi, Eunbeen Lee, Jeongeun Cho, and Seongmin Kim*

Department of Convergence Security Engineering, Sungshin Women's University, Seoul, Korea
{20190886, 20200965, 20200947, 20200958, sm.kim}@sungshin.ac.kr

Abstract

The exponential growth of the Internet of Things (IoT) has made these technologies integral to our daily lives, but it also raises significant security concerns as IoT and mobile devices become prime targets for cyber threats from the Dark Web. Detecting Darknet traffic is crucial in the IoT environment to prevent data breaches and potential large-scale cyberattacks on interconnected devices. Recent studies have improved Darknet traffic detection using supervised machine learning (ML) models. However, existing Darknet traffic datasets are both limited and outdated to address up-to-date cyber threats. We introduce a novel methodology using contrastive learning for Darknet traffic detection, an approach that effectively learns from both similar and dissimilar data pairs. This method addresses the limitations of traditional ML models that rely heavily on labeled data and establishes an Intrusion Detection System (IDS) capable of adapting to modern cyber threats. Our model, tested on the Darknet traffic dataset, achieved a 92% accuracy, indicating that it adeptly learned the distinct characteristics between Darknet and benign traffic.

Keywords: Cybersecurity, Contrastive learning, Internet of Things (IoT), IDS system

1 Introduction

The Internet of Things (IoT) technology has grown exponentially and become an indispensable element in our daily lives. However, such innovation opens up security concerns as IoT and mobile devices both become a target of cyber threats from the Dark Web and harness for conducting illicit behaviors[1]. Distinguishing darknet traffic is vital because undetected malicious activities can compromise vast numbers of interconnected devices, leading to data breaches and potential large-scale cyberattacks on critical infrastructures. To address this issue, recent studies have conducted supervised ML models to detect Darknet traffic and have improved detection accuracy[2]. However, existing Darknet traffic datasets are both limited and outdated to address up-to-date cyber threats.

In this study, we propose a novel Darknet traffic detection methodology by leveraging contrastive learning. Contrastive learning utilizes positive pairs (similar data) and negative pairs (different data) to learn the similarities and differences between data and trains the model through a loss function[3]. Typically, it is designed to cover the limitations of traditional ML models dependent on labeled samples, thereby establishing an Intrusion Detection System (IDS) capable of adaptive responses to modern cyber threats. Upon training and testing the proposed model using the Darknet traffic dataset, our proposed model delivers 92% accuracy, indicating that the model adeptly learned the distinct characteristics between Darknet and benign traffic.

The 7th International Conference on Mobile Internet Security (MobiSec'23), Dec. 19-21, 2023, Okinawa, Japan Article No. W-7
*Corresponding author: Department of Convergence Security Engineering, Sungshin Women's University, Seoul, 02844, Republic of Korea, Tel: +82-02-920-7449

2 Background

2.1 Darknet

The Darknet, commonly known as Darkweb or Deepweb is a network where anonymity is preserved for both client and server IP addresses. This vast segment of the internet is larger than the surface web[4], not indexed by traditional search engines, and requires specialized tools or applications for access. These instruments conceal the IP address of users and encrypt communications, thereby impeding the efforts of monitoring and tracking. However, the veil of anonymity provided by the Darknet can be exploited as a refuge for illicit transactions and criminal enterprises, frequently functioning as a conduit for illegal drug trade, unauthorized arms sales, and the distribution of hacking utilities. In particular, the Tor network[5], which represents a substantial segment of the Darknet, utilizes onion routing technology to transmit data through multiple layers of encryption, ensuring that no individual relay can simultaneously identify the source and the destination, thus safeguarding user anonymity.

2.2 Contrastive learning

Contrastive learning is a technique in unsupervised machine learning, that focuses on distinguishing similar and dissimilar data pairs. This approach helps models learn robust features useful in classification and recognition tasks, particularly beneficial where labeled data is limited. The Siamese networks[6], which utilizes the concept of contrastive learning, is a neural network architecture with two parallel, identical subnetworks sharing the same parameters. This design allows simultaneous processing of two inputs for comparison. They excel in learning distinct features, making them ideal for tasks like similarity detection, face recognition, and signature verification.

3 Proposed Method

This section details the procedure of processing a Darknet traffic dataset, describing the architecture of a contrastive learning-based classification model, and the steps for training and testing the model.

3.1 Dataset

The Darknet is a hidden section of the internet, inaccessible by standard search engines, and serves as a hub for illegal activities and the distribution of hacking tools. To detect Darknet traffic targeting IoT devices, our model utilizes the CIC-Darknet2020 dataset[7], which is comprised of VPN and Tor network traffic integrating two previously released datasets: ISCXTor2016 and ISCXVPN2016.

3.2 Model training

We transformed the 64 features extracted from the dataset into an 8x8 tensor, creating 10,000 pairs of traffic data, and used this as the input to the model. The base network learns the features from data labeled as Darknet and Benign respectively, while the Siamese model learns features through similar and dissimilar pairs of traffic. For the training of the Siamese model, we labeled traffic with the same label as a similar group and traffic with different labels as a dissimilar group.

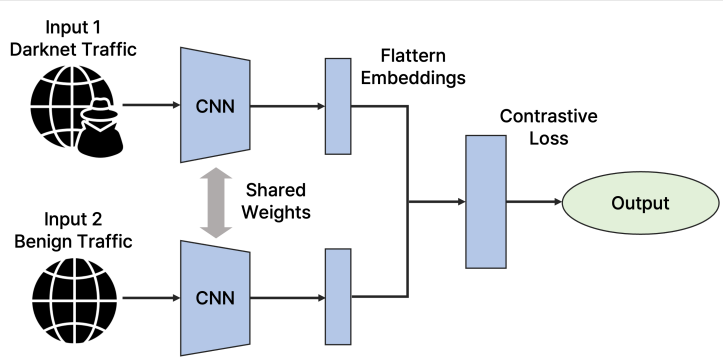

Figure 1: Model Structure

The base network, a simple CNN, is embedded within the Siamese networks. They are connected in such a way that the base network's feature extraction capabilities are utilized to compare pairs of inputs in the Siamese networks. Siamese networks are trained using contrastive loss by determining how close or distant pairs of samples are in the embedding space. For pairs with the same label, the loss aims to minimize their distance, while for pairs with different labels, it aims to ensure a minimum margin between them. The total loss is the sum of the losses for similar and dissimilar pairs.

The contrastive loss is computed according to equation (1), where y is the label indicating whether the pair is similar (1 for dissimilar pairs, 0 for similar pairs). D denotes the distance between the embeddings of the paired samples and m is a margin, which is a hyperparameter that defines how far apart the embeddings of dissimilar pairs should be. Note that the margin value is empirically set to 10.

$$L(W) = \frac{1}{2N} \sum_{i=1}^{N} yD^2 + (1-y)\max(0, m-D)^2, \quad D = \|f_W(x_1) - f_W(x_2)\| \tag{1}$$

4 Evaluation

The final model utilizes the contrastive loss function, aiming to minimize the distance for similar pairs and maximize it for dissimilar ones. To calculate the detection accuracy, the model divides the prediction values into 0 and 1 based on a preset threshold and contrasts it with the actual labels(similar and dissimilar). We utilize the Adam optimizer with a learning rate of 0.001. Both model was trained with a batch size of 64 over 10 epochs. 20% of the training data was reserved for validation.

The proposed model achieved an accuracy of 92%, which is effective in discriminating the Darknet traffic. Contrastive learning is a method that can enhance learning performance in situations where labeled data is scarce, but unlabeled data is abundant. Given that it is challenging to obtain labeled data and there's a lack of up-to-date public datasets for Darknet traffic, contrastive learning can be a promising solution for the darknet traffic problem. To fully leverage the advantages of contrastive learning, incorporating techniques such as data augmentation can potentially maximize learning efficacy. Such an approach is anticipated to be beneficial in the rapidly evolving IoT technological landscape, especially in the extraction of

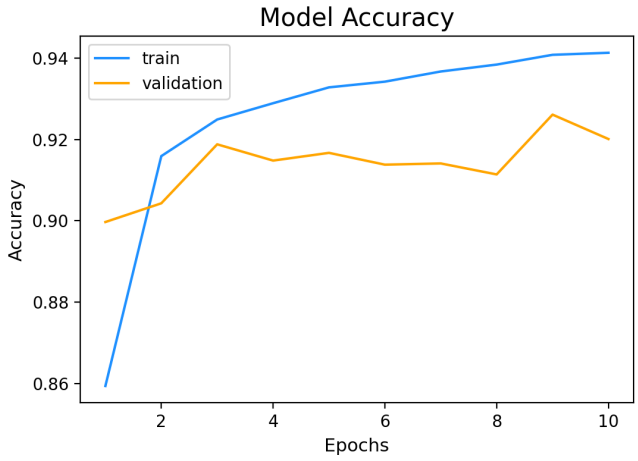

Figure 2: Model Accuracy

features from the most recent traffic data.

5 Conclusion

In this study, we presented a novel approach to detecting Darknet traffic through contrastive learning, tailored for the evolving threats in the IoT landscape. Our model, leveraging the CIC-Darknet2020 dataset, achieved a notable 92% accuracy, demonstrating its effectiveness in distinguishing between Darknet and benign traffic. This success is primarily due to the model's ability to learn from both similar and dissimilar data pairs, overcoming the limitations of traditional machine learning models that rely heavily on labeled data. The use of contrastive learning in this context is particularly innovative, addressing the scarcity of up-to-date, labeled Darknet traffic data. The results indicate that our approach could be pivotal in enhancing cybersecurity measures against Dark Web threats, making it a significant contribution to the field of IoT security.

Acknowledgments

This work was supported by the Sungshin Women's University Research Grant of H20230050. 5

References

[1] Liangmin Wang, Hantao Mei, and Victor S. Sheng. Multilevel identification and classification analysis of tor on mobile and pc platforms. *IEEE Transactions on Industrial Informatics*, 17(2):1079–1088, 2021.

[2] Qasem Abu Al-Haija, Moez Krichen, and Wejdan Abu Elhaija. Machine-learning-based darknet traffic detection system for iot applications. *Electronics*, 11(4), 2022.

[3] Pranjal Kumar, Piyush Rawat, and Siddhartha Chauhan. Contrastive self-supervised learning: review, progress, challenges and future research directions. *International Journal of Multimedia Information Retrieval*, 11(4):461–488, 2022.

[4] M Bergman. The deep web: Surfacing hidden value. *Taking License*, 7, Aug 2001. Taking License.

[5] Tor Project. www.torproject.org. Accessed: 2023-12-16.

[6] Richard Zemel Gregory Koch and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2, 2015.

[7] Lazaros Alexios Iliadis and Theodoros Kaifas. Darknet traffic classification using machine learning techniques. In *2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, pages 1–4, 2021.