Examining the Security of DDoS Detection Systems in Software Defined Networks

Ahmed Abusnaina
University of Central Florida
ahmed.abusnaina@knights.ucf.edu

DaeHun Nyang
Inha University
nyang@inha.ac.kr

Murat Yuksel
University of Central Florida
Murat.Yuksel@ucf.edu

Aziz Mohaisen
University of Central Florida
mohaisen@ucf.edu

ABSTRACT

With the rapid development of Software-Defined Networking (SDN) advocating a centralized view of networks, efficient and reliable Distributed Denial of Service (DDoS) defenses are necessary to protect the centralized SDN controller. In this work, we explore the robustness of DL-based DDoS defenses in SDN against adversarial learning attacks. First, we investigate generic off-the-shelf adversarial attacks to test the robustness of DDoS defenses in SDN. Then, we propose Flow-Merge for realistic adversarial flows while achieving a high evasion rate. The evaluation shows that the proposed Flow-Merge is able to force the DL-based DDoS defenses to misclassify 100% of benign flows as malicious.

CCS CONCEPTS

• **Security and privacy** → **Intrusion detection systems**; **Denial-of-service attacks**.

KEYWORDS

Distributed Denial of Service, Intrusion Detection, Deep Learning, Adversarial Attacks

ACM Reference Format:
Ahmed Abusnaina, DaeHun Nyang, Murat Yuksel, and Aziz Mohaisen. 2019. Examining the Security of DDoS Detection Systems in Software Defined Networks. In *The 15th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '19 Companion), December 9–12, 2019, Orlando, FL, USA.* ACM, Orlando, FL, USA, 2 pages. https://doi.org/10.1145/3360468.3368174

1 INTRODUCTION

Software Defined Networking (SDN) overcomes scalability challenges in network management by a centralized view of a network with many components. A programmable controller in SDN can see all switches and endpoints in a network and manage flows between them, providing a better and easier network monitoring and enhancing security compared to traditional networks [7]. On the other hand, it has been shown that SDNs are vulnerable to Distributed Denial of Service (DDoS) attacks, which target the centralized controller [3]. Those attacks flood services with malicious or undesirable traffic, disallowing them from processing legitimate requests. Moreover, adversarial attacks are on the rise, an adversary can make the model misclassify by applying small perturbations to the input, resulting in adversarial examples (AEs) [1, 4, 6]. The crafted AEs are very similar to the original ones, and are not necessarily outside of the training data manifold, making them hard to be distinguished from legitimate ones. In the context of IDSs in SDN, the failure of the anomaly detector may result in disastrous consequences, such as the failure of the entire network since a successful DDoS on the centralized controller effectively brings all the switches down.

Contributions. 1) We investigated applying generic adversarial learning methods on DL-based IDS in SDN. Our experiments demonstrate that although these methods can achieve high evasion rate, the generated adversarial flows are not realistic, precluding the applicability of generic approaches to DL-based IDSs in SDN. 2) We propose Flow-Merge, an approach specifically designed to fool DL-based IDSs in SDN, while maintaining the characteristics of original flows. Flow-Merge is able to achieve targeted and untargeted attacks.

2 APPROACH

Generic Adversarial Attacks. Generic adversarial attacks were developed for image misclassification by small perturbation to the input, leading to incorrect model output. In this study, we utilize five adversarial attack algorithms: Carlini & Wanger (C&W), Elastic-Net, DeepFool, Momentum Iterative Method (MIM), and Projected Gradient Decent (PGD) to generate AEs by applying a small perturbation to the input, leading to incorrect model output. Although the aforementioned methods excel on images, they were not designed to consider feature dependencies. These methods focus on misclassification, regardless of the functionality, *i.e.,* a malicious adversarial flow may be misclassified as benign with a feature space representing zero packets, resulting in functionality preserving issues.

Flow-Merge. In this approach, the features of the original and the mask flows are combined using one of two approaches: accumulating or averaging. The count-based features, such as the number of incoming/outgoing TCP flows, are accumulated, while the ratio-based features, such as the fraction of TCP flows over the total number of incoming/outgoing flows, are averaged in a weighted form. The weights for each ratio-based feature are protocol-specific. For instance, the number of incoming TCP flows determines the weight of the fraction of TCP flows with the SYN flag set. At the feature level, let $X = \{x_1, \ldots, x_k\}$ be the features of the original

CoNEXT '19 Companion, December 9–12,2019, Orlando, FL, USA

ACM ISBN 978-1-4503-7006-6/19/12.
https://doi.org/10.1145/3360468.3368174

Table 1: Per-class flow records distribution.

Class types		# of records Train	Test
Normal (C0)		49,179	21,076
DDoS attacks	TCP (C1)	5,471	2,344
	UDP (C2)	5,273	2,260
	ICMP (C3)	1,602	686
	TCP & UDP (C4)	4,694	2,011
	TCP & ICMP (C5)	4,739	2,031
	UDP & ICMP (C6)	4,437	1,902
	All (TCP & UDP & ICMP) (C7)	5,615	2,407
Total		81,010	34,717

Table 2: Misclassification rate against adversarial attacks. Acc is the accuracy of the baseline model, EN is ElasticNet method, D. Conf. refers to the detection model., and C. Conf. refers to the classification model.

Model	Baseline Model(%)			Attack Misclassification Rate (%)				
	Acc	FNR	FPR	C&W	EN	DF	MIM	PGD
D. Conf.	99.83	0.05	0.34	98.92	99.84	62.78	93.65	99.76
C. Conf.	96.05	8.18	0.54	95.24	97.13	47.14	88.57	99.26

flow, and $Y = \{y_1, \ldots, y_k\}$ be the features of the mask flow. A feature vector of the masked flow (modified one) is then calculated as $Z = \{z_1, \ldots, z_k\}$ such that $z_i = [n/(n+m)]x_i + [m/(n+m)]y_i$, where n is the number of packets associated with the studied protocol in the first flow, and m is the number of packets associated with the same protocol in the second flow. The count-based features are simply accumulated (i.e., $z_i = x_i + y_i$).

3 EVALUATION

Dataset. For benchmarking, we use the dataset provided by Niyaz *et al.* [5] to evaluate the performance of the methods in this study. The dataset consists of 68 statistical features, divided into TCP-related features (34), UDP-related features (20), and ICMP-related features (14). The per-class distribution of the dataset is shown in table 1.

Intrusion Detection System. We trained two CNN-based models for detection (two-class classification) and classification (eight-classes classification) of intrusion attacks. The detection and classification models achieved a remarkable accuracy rates of 99.83% and 96.05%, respectively.

Adversarial Attacks Configuration. To investigate the robustness of the utilized models against adversarial attacks, we generate AEs based on five different off-the-shelf methods: C&W, ElasticNet, DeepFool, MIM, and PGD. The misclassification results are reported in table 2, we achieved an overall misclassification rate of up to 99.84% using ElasticNet.

Flow-Merge. The experiments are carried out for both detection and classification. For detection, the goal is to misclassify the malicious flow into benign and vice versa. Table 3 shows the detailed results of misdetection rates using Flow-Merge for each dominant flow. The results show that the adversary can forge a malicious flow classified as a benign flow for all three dominant flows. For misclassification, we achieved a 100% misclassification rate from all attack classes to benign. The detailed results of benign targeted misclassification attack is provided in Table 4.

Table 3: Flow-Merge misdetection rates. Columns and rows are the original and predicted labels.

Flow type		Benign	Malicious
TCP	Benign	0.003	0.986
	Malicious	0.997	0.014
UDP	Benign	0.888	0.818
	Malicious	0.112	0.182
ICMP	Benign	0.892	0.939
	Malicious	0.108	0.061

Table 4: Misclassification to benign using TCP, UDP, and ICMP dominant flows. Columns refer to the original label and rows refer to the predicted classes.

Flow type		C0	C1	C2	C3	C4	C5	C6	C7
TCP	C0	1	1	1	1	1	1	1	1
	C1–C7	0	0	0	0	0	0	0	0
UDP	C0	1	1	1	1	0.989	0.996	1	0.999
	C1	0	0	0	0	0	0	0	0
	C2	0	0	0	0	0.004	0	0	0.001
	C3	0	0	0	0	0.004	0	0	0
	C4–C5	0	0	0	0	0	0	0	0
	C6	0	0	0	0	0.001	0.004	0	0
	C7	0	0	0	0	0.002	0	0	0
ICMP	C0	1	1	1	1	0.999	1	1	1
	C1–C6	0	0	0	0	0	0	0	0
	C7	0	0	0	0	0.001	0	0	0

4 CONCLUSION

This work investigated the robustness of DL-based DDoS defenses in SDN against adversarial attacks. Flow-Merge utilizes a weighted merging technique over ratio-based features to craft the AEs. The evaluation results show a misclassification rate of 99.84% using generic adversarial attacks. Moreover, Flow-Merge produces realistic adversarial flows for targeted misclassification with a success rate of 100%, misclassifying all malicious flows into benign. The extended version of this work can be found in [2].

Acknowledgement. This work is supported in part by NVIDIA GPU Grant, NRF-2016K1A1A2912757, and NSF awards 1647189, 1814086, and 1643207.

REFERENCES

[1] Ahmed Abusnaina, Aminollah Khormali, Hisham Alasmary, Jeman Park, Afsah Anwar, and Aziz Mohaisen. 2019. Adversarial Learning Attacks on Graph-based IoT Malware Detection Systems. In *Proceedings of the 39th IEEE International Conference on Distributed Computing Systems, ICDCS*.

[2] Ahmed Abusnaina, Aminollah Khormali, DaeHun Nyang, Murat Yuksel, and Aziz Mohaisen. 2019. Examining the Robustness of Learning-Based DDoS Detection in Software Defined Networks. In *Proceedings of the IEEE Conference on Dependable and Secure Computing (IDSC 2019)*.

[3] K. Kalkan, G. Gur, and F. Alagoz. 2017. Defense Mechanisms against DDoS Attacks in SDN Environment. *IEEE Communications Magazine* 55, 9 (September 2017), 175–179.

[4] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. 2016. DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. 2574–2582.

[5] Quamar Niyaz, Weiqing Sun, and Ahmad Y. Javaid. 2017. A Deep Learning Based DDoS Detection System in Software-Defined Networking (SDN). *ICST Transactions on Security and Safety* (2017).

[6] Nicolas Papernot, Patrick D. McDaniel, Somesh Jha, Matt Fredrikson, Z. Berkay Celik, and Ananthram Swami. 2016. The Limitations of Deep Learning in Adversarial Settings. In *Proceedings of the IEEE European Symposium on Security and Privacy*. 372–387.

[7] Lisa Schehlmann, Sebastian Abt, and Harald Baier. 2014. Blessing or curse? Revisiting security aspects of Software-Defined Networking. In *Proceedings of the Network and Service Management*. 382–387.