# Denial of Service Attacks: Detecting the Frailties of Machine Learning Algorithms in the Classification Process

Ivo Frazão[1], Pedro Henriques Abreu[1], Tiago Cruz[1(✉)], Hélder Araújo[2], and Paulo Simões[1]

[1] Centre of Informatics and Systems, Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal
{icosteira,pha,tjcruz,psimoes}@dei.uc.pt
[2] Institute for Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, Coimbra, Portugal
helder@isr.uc.pt

**Abstract.** Denial of Service attacks, which have become commonplace on the Information and Communications Technologies domain, constitute a class of threats whose main objective is to degrade or disable a service or functionality on a target. The increasing reliance of Cyber-Physical Systems upon these technologies, together with their progressive interconnection with other infrastructure and/or organizational domains, has contributed to increase their exposure to these attacks, with potentially catastrophic consequences. Despite the potential impact of such attacks, the lack of generality regarding the related works in the attack prevention and detection fields has prevented its application in real-world scenarios. This paper aims at reducing that effect by analyzing the behavior of classification algorithms with different dataset characteristics.

**Keywords:** Denial of Service attacks · Intrusion detection systems
Classifier performance

## 1 Introduction

Cyber-Physical System (CPSs) play an important role in today's society, particularly in the control of Critical Infrastructure (CIs), whose uninterrupted operation is essential for the safety and livelihood of a modern society [1].

A Denial-of-Service (DoS) attack is an attack on the quality and/or availability of a service that aims to disrupt the normal operation of an infrastructure by preventing or degrading the communication between its components. A Distributed DoS (DDoS) attack is a variant which further complicates its detection and prevention since it doesn't (or doesn't appear to) originate from a single source, making it very difficult to distinguish legitimate and illegitimate network traffic. Despite the relative abundance of published work in the field of intrusion detection devoted to DoS attacks, its bulk is mostly focused on the


E. Luiijf et al. (Eds.): CRITIS 2018, LNCS 11260, pp. 230–235, 2019.
https://doi.org/10.1007/978-3-030-05849-4_19

perspective of novel algorithms or domain-adaptation of known algorithms, preventing a proper generalization of their results and real-world applications. This article constitutes the first step to providing an insight at the frailties of ML algorithms in the classification process of DoS attacks, specifically by evaluating the impact that the size of the datasets and the relative scale of the attack traces within such datasets affects the performance of common algorithms. While performance comparisons are already available, they don't provide insights about how specific dataset characteristics relate to the performance obtained.

## 2 Related Work

Studies about the topic of DoS attack detection within the scope of common Internet infrastructures are abundant, such as the work in [3] (2003), where the authors performed a comparison between a comprehensive set of ML algorithms against the KDD dataset, which demonstrated that the classifiers were not capable of detecting all the attacks with high success, but, by using the best classifier for each type of attacks, a multi-classifier could be built, which outperformed every algorithm; or in [4] (2016), the authors proposed an ensemble-based multi-filter feature selection method for DDoS detection.

Domain-specific intrusion detection strategies for CPSs have also been proposed, encompassing diverse approaches to the subject, e.g., by implementing techniques derived from common defense mechanisms for Internet-exposed or ICT networks, such as the work in [5] (2005), where the author proposed an anomaly-based IDS, which makes predictions based in historical, exemplar observations of the traffic used for weighted distance calculation; or by modeling the system's behavior, as the work made in [6] (2013), where the authors presented a Deterministic Finite Automata to model the network, or in [7] (2014), where the authors use a variable-order Markov chain to determine the state of the system and detect anomalous occurrences, or even in [8] (2015), where the authors present a finite-state machine modeling technique for each type of register in a PLC to detect anomalous variances of the values. The effects of DoS attacks on these systems have been evaluated in works such as [9], where the authors analyzed the impact of a DDoS attack on the state of a simulated SCADA server, or in [10], where the authors simulated IP packet, TCP SYN, and IEC 60870-5-104 APCI packet flooding against an RTU and analyzed their impacts on the availability of the system. Such works are important to identify the relevant traffic features that can provide evidence about the existence of attacks, as well as the negative impact at a system-wide level, and not only on the component(s) directly affected by the attack. In [11] (2014), the authors performed an evaluation of the classification performance of multiple ML methods in order to explore the aptitude of such techniques for the detection of disturbances in the electrical grid, implementing three testing schemes, using multi-, three-, and binary-class classification of events, in 15 datasets. The first noteworthy result from this work showcased the consistency of the results for each learner, regardless of the dataset being used. Nevertheless, there can be no

definitive conclusion on using different classification schemes, since each learner had a different response to each scheme.

## 3 Experimental Setup

The availability of datasets or network traces containing normal SCADA operations, as well as attacks aimed at those systems, is very limited. To overcome this limitation a testbed was used to generate those datasets, which emulates a CPS process controlled by a SCADA system using the MODBUS protocol. It consists of a liquid pump simulated by an electric motor controlled by a VFD, which in its turn is controlled by a PLC. The motor speed is determined by a set of predefined liquid temperature thresholds, whose measurement is provided by a MODBUS RTU device providing a temperature gauge, simulated by a potentiometer connected to an Arduino. The PLC communicates with the HMI controlling the system and horizontally with the RTU, providing insightful knowledge of how this type of communications may affect the overall system.

There are several types of DoS attacks that are effective against SCADA-based systems using the MODBUS over TCP protocol. For analysis purposes, a subset of the possible attacks was implemented in the testbed, namely ping flooding, TCP SYN flooding, and MODBUS Query flooding - Read Holding Registers, which targeted the PLC. While the first two attacks attempt to overwhelm the capacity of the network or the networking subsystem in the target device with requests (operating mostly at OSI layers 2 to 4), the third attack works at the SCADA protocol layer, flooding the device with read request operations which may lead to side effects such as device resource exhaustion, scan cycle latency deviations or loss of connectivity. The first two attacks were implemented using the *hping3* tool, using its ability to spoof the packet's IP address, whilst the last attack was implemented using an adaptation of the SMOD tool.

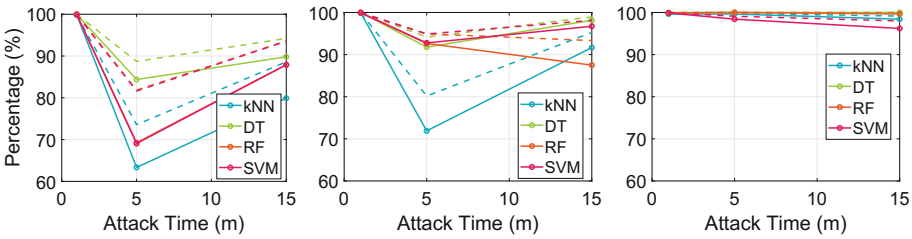
Concerning the aim of this work - to analyze the impact of the dataset size and the relative magnitude of the attack traces within such datasets on the behavior of machine-learned classifiers for DoS attacks - the following experiments were made: varying the time of the capture (30 min and 1 h of capture); and varying the time of the attack (1, 5, and 15 min of attack within each capture). The network captures were acquired from the testbed using the *tshark* network analyzer tool – for this purpose, the network switch that was used to interconnect the equipment was configured with a mirror port. This capture was then processed for feature extraction within Matlab, where a total of 68 features were extracted (packet timestamps, inter-packet arrival times, binary features defining which protocols were involved, and every field of the Ethernet, ARP, IP, ICMP, UDP, TCP and MODBUS over TCP headers). However, to generalize the captures for normal use of the system, the timestamps were removed from the datasets before the analysis.

Four of the most used classifiers were implemented in this analysis study, namely: k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), resorting to the Matlab implementations

of the algorithms. In order to validate the models created, a cross-validation procedure was performed with a 70%/30% ratio for training and validation sets.

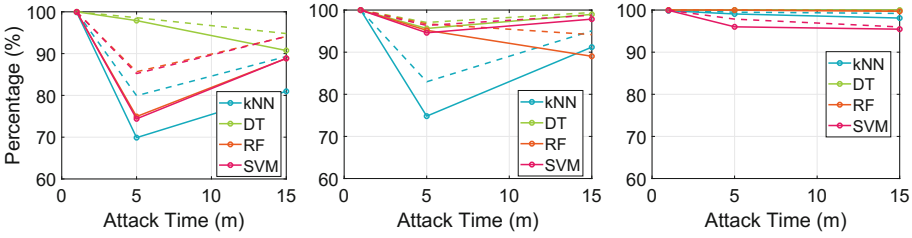
## 4 Results and Discussion

The results for the implemented algorithms are shown in Figs. 1 and 2 (full lines represent the accuracy, whilst dashed lines represent the F1-scores). The DoS attacks that were implemented are labeled as follows: (1) Ping DDoS Flood; (2) TCP SYN DDoS Flood; and, (3) MODBUS Query Flood.




(a) Attack 1. (b) Attack 2. (c) Attack 3.

**Fig. 1.** Results of the implemented classifiers with 0.5 h of capture.




(a) Attack 1. (b) Attack 2. (c) Attack 3.

**Fig. 2.** Results of the implemented classifiers with 1 h of capture.

On a first approach, the analysis of the results reveals a trend for consistent high accuracy results for the smallest attack timespan. However, this is a misleading result, given the high imbalance in the data present in these situations, associated with the small variance of the attacks in such a small timespan. When the attacks increase in size, a decay of the performance can be observed, explained by the decrease of the aforementioned effects. The increasing attack timespan overcomes the imbalance within the dataset, allowing the algorithms to systematically learn more differences between normal and anomalous traces

and improve the results – a similar effect is also accomplished with an increase in the size of the capture. Although the increase of the size of the capture increases the imbalance of the dataset, it also increases the number of packets available to differentiate the traces, improving the performance of the classifiers, as can be seen by the reduced decay in performance when the time of capture increases. The improvement of the results as a consequence of the increase in the relative size of the attack traffic is also reduced when the size of the capture increases.

The DT classifier presents the best results of all the studied classifiers. By analyzing the trees obtained, some of the attacks were classified by detection of the reduced inter-packet arrival times (a good metric for flooding attacks), however, when the inter-packet arrival times were not sufficient, the algorithm tended to overfit the data and, consequentially, exhibited higher performances but lacked generality. The RF algorithm aims to prevent these overfitting issues and, consequentially, showcases worse accuracy.

All the classifiers presented unusually good results for the third attack (the MODBUS query flood), prompting a deeper analysis of the dataset obtained. This allowed for the detection of a field with little variance during the attack, which allowed the algorithms to detect the attack with ease. Consequentially, inferring upon these results may be overreaching, requiring further analysis and an adaptation of the implemented attack in future works.

In conclusion, the effects of data imbalance on the classification process constitute a real problem that can lead to unintentional misleading when analyzing classifier performance. Moreover, this situation can be hard to detect since the datasets used for CPS security research are frequently restricted (and consequentially, difficult to analyze and characterize).

## 5 Conclusion and Future Work

This article constitutes the first step to providing an insight on the frailties of machine learning algorithms which may lead to proper generalization of the techniques and, consequentially, real-world application. The effects of varying the attack and the capture timespans, when using different algorithms and attacks were studied. It was inferred that, although small attack timeframes provide apparently good results, the classification accuracy starts to decrease as they grow in size and the imbalance of data starts to diminish. Once the data imbalance is overcome, the results start improving again. The overfitting problem is also detected and discussed.

As future work, the authors plan to further pursue this analysis effort, increasing the capture and attack timespans and also diversifying the types of implemented attacks. Finally, future developments of this work will also involve an analysis of how the feature selection process may affect both the time required to create the models for detection and the resulting classification performance.

**Acknowledgements.** This work was supported by the ATENA European H2020 Project (H2020-DS-2015-1 Project 700581).

## References


1. Humayed, A., Lin, J., Li, F., Luo, B.: Cyber-physical systems security: a survey. IEEE Internet Things J. **4**(6), 1802–1831 (2017). https://doi.org/10.1109/JIOT.2017.2703172
2. Zargar, S.T., Joshi, J., Tipper, D.: A survey of defense mechanisms against distributed denial of service (DDOS) flooding attacks. IEEE Commun. Surv. Tutor. **15**(4), 2046–2069 (2013). https://doi.org/10.1109/SURV.2013.031413.00127
3. Sabhnani, M., Serpen, G., More, K.K.: Application of machine learning algorithms to KDD intrusion detection dataset within misuse detection context. In: Proceedings of International Conference on Machine Learning: Models, Technologies, and Applications (MLMTA), January 2003, pp. 209–215 (2003). http://dl.acm.org/citation.cfm?id=1293805.1293811
4. Osanaiye, O., Cai, H., Choo, K.K.R., Dehghantanha, A., Xu, Z., Dlodlo, M.: Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. Eurasip J. Wirel. Commun. Netw. **2016**(1), 130 (2016). https://doi.org/10.1186/s13638-016-0623-3
5. Su, M.Y.: Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers. Expert Syst. Appl. **38**(4), 3492–3498 (2011). https://doi.org/10.1016/j.eswa.2010.08.137
6. Goldenberg, N., Wool, A.: Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems. Int. J. Crit. Infrastruct. Prot. **6**(2), 63–75 (2013). https://doi.org/10.1016/j.ijcip.2013.05.001
7. Yoon, M., Ciocarlie, G.F.: Communication pattern monitoring : improving the utility of anomaly detection for industrial control systems. SENT **14**(February), 110 (2014). https://doi.org/10.14722/sent.2014.23012
8. Erez, N., Wool, A.: Control variable classification, modeling and anomaly detection in Modbus/TCP SCADA systems. Int. J. Crit. Infrastruct. Prot. **10**, 59–70 (2015). https://doi.org/10.1016/j.ijcip.2015.05.001
9. Markovic-Petrovic, J.D., Stojanovic, M.D.: Analysis of SCADA system vulnerabilities to DDoS attacks. In: 11th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, TELSIKS 2013, vol. 2, pp. 591–594 (2013). https://doi.org/10.1109/TELSKS.2013.6704448
10. Kalluri, R., Mahendra, L., Kumar, R.K.S., Prasad, G.L.G.: Simulation and impact analysis of denial-of-service attacks on power SCADA. In: National Power Systems Conference, NPSC 2016, vol. 1 (2017). https://doi.org/10.1109/NPSC.2016.7858908
11. Hink, R.C.B., Beaver, J.M., Buckner, M.A., Morris, T., Adhikari, U., Pan, S.: Machine learning for power system disturbance and cyber-attack discrimination. In: 7th International Symposium on Resilient Control Systems, ISRCS 2014 (2014). https://doi.org/10.1109/ISRCS.2014.6900095