Proceedings of the 33rd International Conference on Information Technologies (InfoTech-2019)
19-20 September 2019, Bulgaria

HIERARCHICAL MACHINE LEARNING FOR IoT ANOMALY DETECTION IN SDN[1]

Digest of paper[2]

Perekebode Amangele†, Martin J. Reed†, Mays Al-Naday†, Nikolaos Thomos†, Mateusz Nowak‡

†*University of Essex* ‡*IITIS PAN*
e-mails: †{p.amangele,mjreed,mfhaln,nthomos}@essex.ac.uk ‡mateusz@iitis.pl
†*UK,* ‡*Poland*

Abstract: The Internet of Things is a fast emerging technology, however, there have been a significant number of security challenges that have hindered its adoption. This work explores the use of machine learning methods for anomaly detection in network traffic of an IoT network that is connected through a Software Defined Network (SDN). The use of SDN allows a hierarchical approach to machine learning with the aim of reducing the packet level processing of anomaly detection at the edge through applying additional centralised machine learning in the SDN controller. For the sake of evaluation, we compare several supervised classification algorithms using a publicly available dataset. The results support a decision-tree based approach and show that the proposed solution promises a considerable reduction in the per-packet processing at the network edge compared to a single stage classifier.
Key words: network security, SDN, anomaly detection, IoT.

1. INTRODUCTION

The Internet of Things (IoT) is a fast growing network of physical devices that will soon encompass billions of devices. However, there have been significant problems with security in IoT devices due to a combination of: the constrained and heterogeneous nature of the end devices themselves; the fact that end-devices are often not visible and lack standard management systems for firmware/software upgrades; and, connecting to the Internet opens them up to attacks both from and to

[1] Perekebode Amangele is sponsored by Petroleum Technology Development Fund of Nigeria. The work was carried out within the project SerIoT, which has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 780139.
[2] The full paper is proposed for including in the IEEE Xplore Digital Library

the Internet. As the IoT devices themselves are often highly constrained, it is not straightforward to operate security services such as intrusion detection/prevention systems on the devices them- selves. Therefore, network based solutions to IoT security are required. Such a network solution is the architecture proposed by the SerIoT project [1], which uses SDN [2] as the network infrastructure to both interconnect the IoT devices and provide the IoT security solution. Our system uses a two-stage classifier to reduce the packet processing effort in an edge classifier by utilising an SDN architecture with an additional centralised anomaly detection layer. Although we address the anomaly detection within the SerIoT SDN architecture, the proposed method is generic and, thus, applicable to other SDN based networks**.** This paper is a summary of a longer paper with the same title and authors available from IEEE Xplore. Here the background is very briefly stated in Section 2 before the proposed solution is given in Section 3. Section 4 describes the classifier design and evaluation before conclusions are finally drawn.

2. BACKGROUND

SDN is a next generation networking concept which offers greater flexibility and control compared to traditional net- works. SDN provides logically centralised control over the net- work and separates the control and data planes, by abstracting the lower-level functionality allowing network management and control to be directly programmable. This is achieved by extracting the network control logic (control plane) from the underlying switches and routers that perform the actual task of traffic forwarding (data plane). With this separation, network nodes become simple, efficient, forwarding devices and the control function is implemented in a centralised controller [2]. The centralised control of SDN allows the SDN controller to deploy network-wide policies for both routing and security in a more flexible and agile manner. The centralised controller can instruct forwarding devices to allow or block traffic (as well as deciding the route). In this work, we will also use SDN to redirect potentially malicious traffic.

3. PROPOSED 2-STAGE HIERARCHICAL MACHINE LEARNING BASED SDN SECURITY SOLUTION

This paper proposes a novel, 2-stage hierarchical Machine Learning process, integrated into an SDN architecture for network traffic anomaly detection and mitigation. We adopt SDN in our framework because of the flexibility it offers, i.e. it has the capacity to dynamically address security requirements of networks, and the ability of a central controller to have a global view of the network. The proposed SDN solution integrates a 2-stage Machine Learning instance as shown in Fig. 1. The first instance of machine learning, i.e., Classifier 1, works on summarised network flow traffic features captured using techniques such as IPFIX [3]. This classifier is implemented in the SDN Controller and serves as a central classifier that works on gross flow level information. This is a reasonable approach as it is inefficient to send every packet of the flow to a central classifier because this could result in a relatively

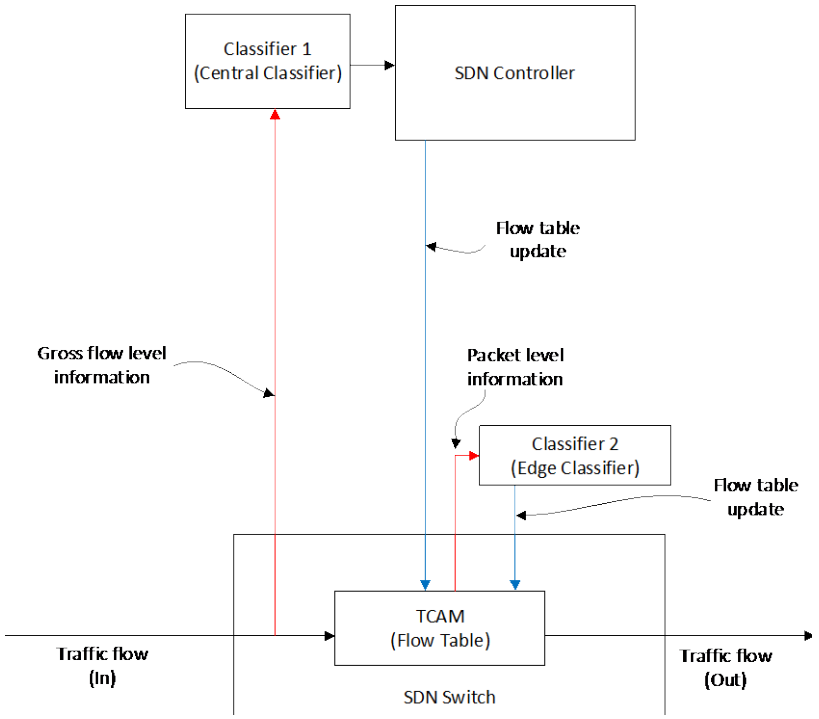


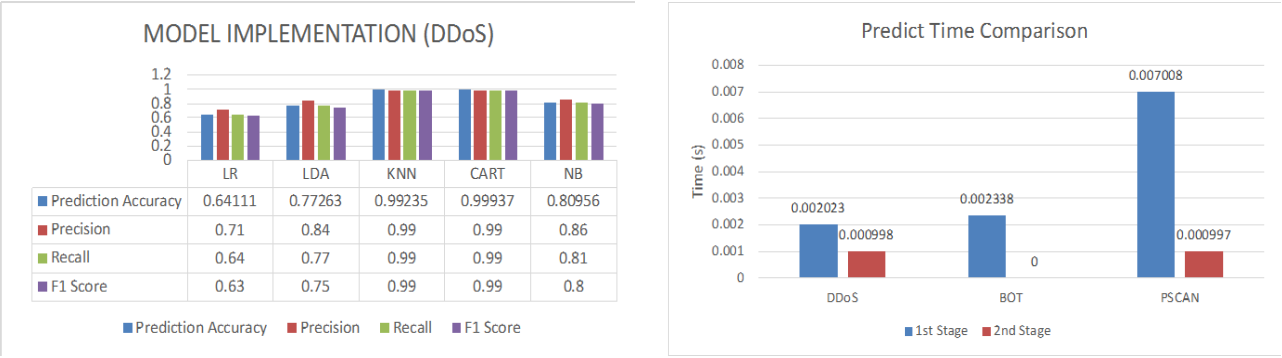
Fig. 1. Hierarchical Machine Learning Architecture for SDN Security (note only one switch of the many is shown)

poorer false positive performance (some good packets may be classified as bad), while the aim of this classifier is to have low false negatives (i.e., to identify all bad packets). The central Classifier 1 is used to identify potentially harmful network traffic which is fed into a second machine learning stage, Classifier 2 as shown in Fig. 1, which works on a per-packet basis at the network edge.

4. CLASSIFIER DESIGN AND EVALUATION

The classifier selection was carried out using the CICIDS2017 dataset which is specifically for network security and intrusion detection testing [4]. The CICIDS2017 dataset is comprised of seven attack categories [4]. Algorithm selection was performed over including both machine learning quality metrics and prediction time. The latter is of concern as the aim is to run the algorithm on real-time network traffic, whereas the model fit time is less important as the fit can be performed offline. Performance was compared across six algorithms from the Scikit-learn Python package [5]: Linear Regression (LR), Linear Discriminant Analysis (LDA), k-Nearest Neighbour (KNN), Classification and Regression Tree (CART), Na¨ive Bayes (NB) and Support Vector Classification (SVC). Fig. 2a shows the summary results except for SVC which was excluded due to slow fit and predict times. CART was selected as it had the best performance across both quality and time metrics. Using the CART prediction, Fig. 2b shows the relative time between classifying only centrally or at the edge in the 2nd stage Classifier 2. This is important as it shows the benefit from moving from an edge-based classifier alone to the proposed two-stage classifier

which leads to a significant reduction in packet level traffic needed to be processed in Classifier 2, thus leading to a more scalable approach for practical deployment.



(a) *(b)*

Fig. 2. a) Performance of different algorithms. b) Predict time of Cassifier 1 vs Classifier 2 using hierarchical approach

5. CONCLUSION

A novel hierarchical Machine Learning architecture for SDN security has been proposed in this paper. The architecture consists of two Classifier stages with the first classifier implemented in the SDN controller and the second implemented at the edge in a processing device co-located with the SDN switch. A range of suitable machine learning algorithms were evaluated, suggesting that a classification and regression tree model is the most suitable algorithm from those investigated. The results show that using the proposed hierarchical approach there is a significant reduction in the number of packets that have to be processed in the classifier associated with the SDN switches.

REFERENCES

[1] E. Gelenbe, J. Domanska, T. Czchorski, A. Drosou, and D. Tzovaras (**2018**). Security for Internet of Things: The SerIoT Project. *Proc. of Int. Symp. on Networks, Computers and Communications (ISNCC)*, Rome, Italy.

[2] D. Kreutz, F. M. V. Ramos, P. Esteves Verissimo, C. Esteve Rothenberg, S. Azodolmolky, and S. Uhlig (**2015**). Software-Defined Networking: A Comprehensive Survey, *Proceedings of the IEEE* vol. 103, No.1, pp. 14-76.

[3] B. Trammell and E. Boschi (**2011**). An introduction to IP flow information export (IPFIX). *IEEE Communications Magazine*, vol. 49, no. 4, pp. 89–95.

[4] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, (**2018**). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. *Proc. of Int. Conf. on Information Systems Security and Privacy (ICISSP*), Funchal, Madeira, Portugal Jan. 2018, pp. 108–116.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vander- plas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duch- esnay (**2011**). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Jan. 2011.