Machine learning and datamining methods for hybrid IoT intrusion detection

Ait Moulay Rachid (rachid_aitmoualy@um5.ac.ma)
Université mohammed V
Abdellatif El Ghazi (abdellatif.elghazi@uir.ac.ma)
International University in Rabat

Research Article

Keywords: IDS, Cloud, Fog Computing, Machine learning, Datamining, Honeypot

Posted Date: May 10th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2895780/v1

Additional Declarations:
Competing interests: The authors declare no competing interests.

Machine learning and datamining methods for hybrid IoT intrusion detection

Abdellatif El Ghazi
TIC Lab, Information and Communication Technology Laboratory
International University in Rabat
Rabat, Morocco
abdellatif.elghazi@uir.ac.ma

Ait Moulay Rachid*
Algebra and Functional Analysis Group
Mohamed V University in Rabat*
Rabat, Morocco
rachid_aitmoualy@um5.ac.ma*

Abstract—**By 2025 Internet of things will reach over 75 billion devices which would exceed number of humans about 8.1 billion. These devices need to be secured from many threats by implementing secure and interoperable solutions in order to guarantee a proper functioning of the infrastructures and systems using the IoT. This is why we proposed a hybrid intrusion detection system installed on the cloud powering another online and real time intrusion detection system on the fog to monitor the communication and detect attacks before it spreads over the network as in the case of Mirai botnet. We will provide details of the different algorithms used to implement this distributed system so as to detect attacks against IoT devices.**

Keywords—IDS, Cloud, Fog Computing, Machine learning, Datamining, Honeypot.

I. Introduction

The use of IoT in different sectors such as (health, Transport, supply management and logistics, smart buildings and homes), also in personal utilities and wearables, they became omnipresent and widespread in many infrastructures and organizations thanks to smart watch, smart TVs, sensors, actuators ...

The internet of things is a new paradigm that connects the physical world (houses, buildings, factories), this new technology can be connected to the internet using sensors to obtain measurements of (temperature, pressure, pollution rate in the area, light, vibration), it may also determine the road condition and help indentifying people in a house or a building using RFID, and actuators that can control these devices using the data collected from sensors.[1,2,3]

To benefit of the full functionalities offered by connected devices that exchange and exploit a vast amount of data with proprietary platforms, which requires the implementation of solutions, capable of securing a heterogeneous network where each sensor or actuator can implement its own protocol especially in recent years, where Internet of things have become the Achilles heels of companies and organizations targeted by malware and viruses that spread using IoT devices because of the lack of standard and insecure protocols. [1,4,31]

We adopted the integration of IoT devices with big data solutions and machine learning algorithms, to analyze and process the data collected by sensors in different fields (smart homes, smart streets…) also in different types of networks (Zigbee, Bluetooth, WIFI) to improve the security and detect malicious devices in an environment that grows and increases in number every day. [5]

By 2025 the number of sensors and actuators can reach 100 billion of devices with a revenue of 3 trillion dollars [6,7,46], which sometimes use web-based protocols called web of things giving the possibility to send data to the cloud servers or even communicate with social networks [8,3].

This rich and uncontrollable environment gives hackers a way to disrupt organizations by using these IoTs in their future attacks, such as DDOS or to make money with them by what we call crypto-mining.

To create an IDS we distinguish 3 types of analytics and training applied:

Anomaly intrusion detector [9,11,14,15]: this system is based on a model that is trained with normal network traffic in order to be able to detect abnormal behaviour and evaluate traffic that is not included in this category (port scan, increase bandwidth usage). It can also disseminate zero day attacks, i.e. never encountered and never known in the training phase, which makes it difficult for attackers to access a network without being detected since this type of system is configured and trained for a network and environment at its normal state. Furthermore, it can be used to generate signatures for misuse intrusion detection systems.

Misuse intrusion detector [9,10,11,12,13] : is a system based only on attack signatures, therefore, it cannot detect zero day attacks and it requires recurrent updates of its knowledge databases since we must first know how the attacks are carried out against a network of IoT devices in order to protect it. One of its abilities is that it is not generating a high rate of false positives and negatives like anomaly based detectors.

Hybrid systems [9]: is a system that combines both an anomaly detector and an abuse detector, in order to create a system that takes advantage of the signatures generated by an anomaly IDS to perform the basic knowledge database update of a misuse detector system to strengthen network security.

However, there are two major categories of intrusion detection systems based either on host or network monitoring :

- Network intrusion detection system (NIDS) : it is installed on the network extremum to detect outside attackers by monitoring and analyzing intercepted packets to spot attack signatures and abnormal behaviour (misuse detection, behaviour analysis) like we stated before. [9,11,32]
- Host Intrusion detection system (HIDS): it is installed on the machine or the host we intend to monitor, its principal purpose is to identify insider threats and detect abnormal

actions like suspicious files or modification of logs, unknown system calls, adding or deleting a user… [9,11,32]

To protect businesses and organizations using IoT, one of the best solutions is integrating an hybrid IDS based on behaviour and misuse detection system against different types of attacks like Dos, MITM, unauthorized access and control from the Internet, private data access, privilege escalation, malware infection ... [4,10,25,26] using a mechanism based on artificial intelligence, statistics, data mining [23] and data collection to minimize false positives and false negatives for unknown or new attacks (zero day) and maximize the detection rate of real intrusions.

The generation of massive data by IoT devices will require using and applying one or more artificial intelligence algorithms to analyze data by coupling it with big data tools such as MapReduce and Hadoop for parallel processing and distributed storage. This article will consist of 4 sections:

- Related works.
- Theoretical Hybrid IDS Framework.
- Metrics and evaluation.
- Conclusion and implementation challenges.

II. RELATED WORKS

To secure IoT devices a framework named EXPOSURE has been proposed to classify domain names into benign or malicious (Botnet, virus link, spam or phishing link), it uses a J48 decision tree which is an implementation of the C4.5 decision tree. The framework is composed of 5 modules the first one is a data collector that looks for malicious and benign domain names using different sources from the internet , a second module that works as a local data collector for monitoring the network where EXPOSURE is installed, a feature analyser that uses both the precedent modules, so that to label domain names accordingly into malicious or benign, then the result of this feature analyser is used by a machine learner module and a classifier module. This framework which was updated every day to keep up with threats that can damage a network it can detect new malicious domain names that weren't in the training set with an accuracy of 98.5% and 0.9% of false positive rate. [20,21,22].

A misuse intrusion detection system was proposed using ID3 algorithm coupled with unsupervised clustering algorithm that process the rules used by snort IDS [19] to feed them to ID3 decision tree in order to facilitate and optimize the classification. This combination of techniques between supervised and unsupervised algorithms have given better results than the naïve processing used by snort which compares an input with the installed signatures. The old technique might become slower if the signatures in the database are very large.

The framework achieved a maximum speed increase of 105% and an average speed of 40.3% and a minimum of 5% [18].

Another IDS implemented a naïve Bayes algorithm which is known to be fast and intuitive [17]. It was tested on three types of attacks like Dos, Scan, Unauthorized Access with an accuracy rate respectively of 99% 96% 90%. This IDS showed better results in term of speed and accuracy but with too many false positives compared to a Neural Network IDS [16].

A lightweight IDS for edge devices using SVM was proposed to detect only DOS attacks, they used the transmission rate of the packet field to train their model because they remarked this attribute was increasing or decreasing depending of the attack's type, or the execution stage (exfiltration of data, malware update). ,

From this attribute they derived three features composed of mean, maximum, median to avoid under-fitting.

They also created multiple features sets to test the performance of the lightweight SVM with three different kernels: Linear kernel, Polynomials, radial basis function. They conducted multiple experiments to choose better parameters for the SVM based IDS, and they have found out that the linear lightweight SVM is much better than other two kernels. Then they compared the performance of this IDS using accuracy and CPU time with other lightweight algorithms like a Genetical based SVM [33] and A-IDS [34] and wfs-IDS [35]. their IDS outperformed other three algorithms in accuracy and CPU time which confirmed its lightweight property because the CPU time is less than the mentioned three algorithms thus it will not consume more energy and resources for an accuracy of 98.3%. The problem with their IDS is that it has not been used on other attacks like remote to local or unauthorized access on the edge node which limited their IDS to detect only DOS, so it can't be generalized on other attack types. [36]

To protect fog nodes and enhance IoT security an adaptive IDS using Artificial Neural Network was created, capable of measuring threats and self-protecting against attacks by closing connection or asking for authentication... depending on the threat level. They used a risk management unit at the end of the output of their model to evaluate the risk of the abnormal behaviour into different levels between 0 (no attack) or secure state to level 3 which is equivalent to a fog node being at destabilized state that can cripple the fog node functionalities. This risk management unit utilizes the output calculated by the Artificial Neural Network model to evaluate the threat activity by verifying the interval of output τ. This unit has the ability to monitor logs to check the authenticity and the periodicity of the actions that has been raised to the risk management unit to measure its threat levels. They trained three models depending on the resources they are trying to monitor, these resources include Memory availability, buffer consumption and CPU usage.

These models showed an error near to zero when they were compared to their real value.

The architecture of ANN is composed of 10 neurons in hidden layer, an activation function using sigmoid symmetric function, and 2 delays unit and one linear output function. For the algorithm optimization they used levenberg marquardt backpropagation algorithm that showed its capability to efficiently distinguish between normal and abnormal activity. The Framework is able to protect against DOS, flooding with accuracy that can reach 97% and precision of 98.4% and recall of 98.9%, with little overhead to the fog node which can be categorized as lightweight because it did not stress the fog node resources. [37]

To maximize detection rate the authors [47] used DNN-KNN algorithm operating on the fog, which implemented a binary classification model, in the first step they used DNN to classify event into malignant or non malignant, it is composed of one input Layer and two hidden Layers, each of them has the same number of neurons, in the hidden Layer they used hyperbolic tangent as activation function, for the output Layer they used two neurons, one neuron for the malignant activity and the other one for the normal behaviour, they used softmax activation function in the output Layer. if one of the neurons does not achieve a defined limit to conclude the classification of the activity as normal or malicious, the suspicious activity will be sent to a feature reduction module implementing Information Gain algorithm which selects the best features for classification, these features are redirected to the KNN algorithm to be finally classified, the result from the k nearest neighbor is considered as final. The DNN-KNN algorithm showed an accuracy rate of 99.77% and recall rate of 99.76% for the NSL-KDD dataset. For the CICDS2017 it also showed a higher accuracy and recall rate attaining 99.85% and 99.87% compared to other implementation using the same datasets.

Another technique to secure edge devices has been proposed, it is composed of 3 modules, the first module is a snort IDS that has the ability to identify and catch malicious devices, this IDS informs a secure load balancer module about the category and the identification of the edge device. This secure load balancer uses a Markov model to confirm the device category (compromised, normal) and calculate its shifting probability, and then it uses another hidden Markov model to decide if the edge device traffic should be diverted to a third module which is a cloud honeypot that aims to monitor and log all the traffic made by the suspected device. The honeypot uses a two stage Markov model to flag an edge device as secure if it has been classified incorrectly as compromised by the precedent module (secure load balancer). This framework can achieve an accuracy of 90%, it has the ability to minimize the false positives by an online honeypot monitoring diverted traffic of suspected edge device to confirm if it has been misclassified, and also It improves the IDS response by updating a database with attacks detected within the network.[38]

III. THEORETICAL HYBRID IDS FRAMEWORK:

A.MOTIVATION

The goal of our hybrid cloud-based and distributed IDS is to minimize false positives and maximize the detection of zero-day attacks for the signature intrusion detection system based on online Incremental SVM [29,39] installed on a fog architecture [28], since it is not able to detect new attacks (zero-day). This signature based IDS will collect signatures from a honeypot [27] installed on the internet, and an anomaly detector system based on Artificial Neural Network [37] that will help enhance detection rate of the online SVM IDS.

B.SYSTEM ARCHITECTURE

The system will be composed of 4 modules, a data collector based on an intelligent honeypot installed on the cloud to detect all known and also zero-day type attacks against IoT devices like sensors, and cameras.

A second component for feature selection and reduction tool that allows to reduce dimensionality and to minimize inputs as well as features in order to make processing and model creation faster and consume less memory, CPU by using PCA coupled with MapReduce to maximize feature reduction speed. A third module based on Artificial Neural Network to detect anomalous behaviour and update a database of the attacks used by the fourth module, which is an online signature based IDS that uses a database of attacks updated by both the online honeypot and the third module (Artificial Neural Network) to help detect new zero day type attacks as depicted in Fig. 1.

1) Data Collector Using Smart Honeypot

An intelligent honeypot [24,27] will be used to collect and store attacks in the server. Then process them in real time in the cloud to automatically update the misuse detector in the network where edge and IoT devices are installed in order to predict new zero day attacks that are not able to disseminate. This system which is inspired from a practical implementation [24] will be composed of a hybrid honeypot capable of extending the time of attackers connections and sessions, so that it goes to the next step to recover the bits of code used for the persistence of attackers and the exploitation of IoT devices to be used for future attacks.

In order to implement this system, we will use machine learning algorithms like the hidden Markov as well as deep learning and reinforcement learning algorithms like Q-learning.

This honeypot will be initially naive but gradually will make updates to its internal knowledge database by searching on the internet for answers wanted and desired by attackers using platforms like shodan, censys.io, zoomeye and also masscan to extend their sessions. These responses collected from the Internet will then be stored in the database to be selected afterwards by learning algorithms, to increase further the session time and retrieve the exploitation code (payload).

Attacks and payloads sent by attackers will be stored for future processing in order to send them to signatures database, so that the malicious detector updates his model for detecting new zero day attacks (unknown).

However, in order to make this process fast for the incremental learning algorithm we will use big data tools like MapReduce and also apply algorithms for feature reduction like PCA to update their knowledge base of their signature. The tools used to collect the data will be tcpdump and Wireshark. [30]

2) Feature Selection and processing Tool

To make the online SVM IDS work faster we need to use principal Component Analysis (PCA) [42] for feature reduction, it's principally used for data optimization and compression. We need also to use MapReduce [41] to process the constant flow and large quantity of data captured by the honeypot and the Artificial Neural Network IDS. We think this combination of PCA and MapReduce will allow the online IDS to be more adaptive and respond swiftly to new attacks without too much delay.

However, to achieve the precedent goal we will use a multicore machine that will calculate every summation expressions within eigenvectors of the covariance matrix Σ used by the PCA :

$$\Sigma = \frac{1}{m}\left(\sum_{i=1}^{m} x_i x_i{}^{T}\right) - \left(\left(\frac{1}{m} * \sum_{i=1}^{m} x_i\right) * \left(\frac{1}{m}\sum_{i=1}^{m} x_i{}^{T}\right)\right) \quad (1)$$

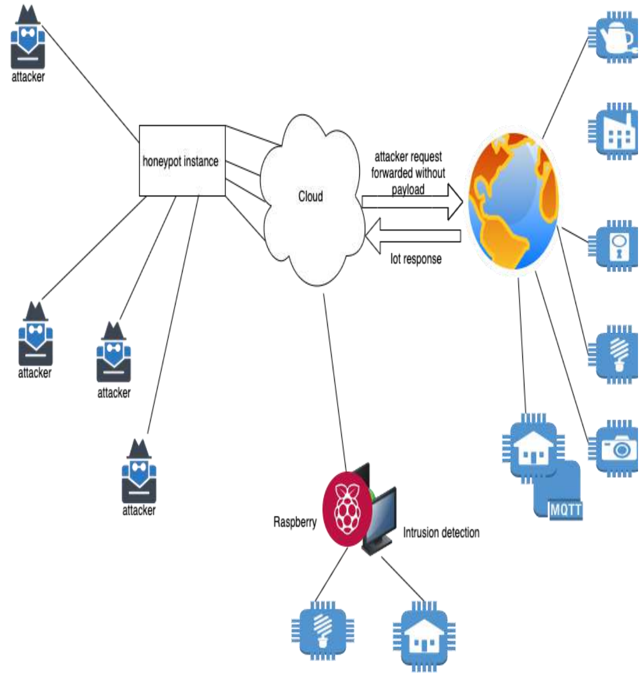

Figure 1 : Framework Architecture

This will allow MapReduce to use each core to calculate summations separately and combine the results to calculate the final covariance matrix Σ [40].

3) *Artificial Neural Network*

To train the Artificial Neural Network (ANN), to help us discover anomalous behaviour we will use a testbed composed of :

- Several smart bulbs.
- Smart TV.
- Temperature sensor.

We chose Multilayer Perceptron (MLP) to train our model from normal and attack data captured by Wireshark. The MLP is composed of one input Layer, one Hidden Layer, and one output Layer as shown in Fig. 2. For the hidden layer we will have 7 neurons, and the function applied for activation is sigmoid function. In the input and hidden layer we will have one bias unit noted as “b”. The MLP will use feed-forward algorithm to calculate signal value “a” of each neuron connection from the input layer to hidden layer using weight $w_i$of an input $x_i$with “n” as the total number of inputs, and then calculate the error in the output layer by comparing the final signal with the expected result. To calculate this output signal of a neuron we apply this equation :

$$a = f(\sum_{i=0}^{n} w_i x_i + b). \qquad (2)$$

After calculation of the error, MLP uses backpropagation algorithm to forward back the error to each layer to recalculate and adjust the weights and bias to minimize the error produced in the feed-forward step.

However, after we train our model offline we will proceed on the online phase where we will install the MLP based IDS on a cloud server to validate traffic classified by the online SVM IDS as normal to make sure no new attacks have passed undetected. Otherwise, IDS behaviour will update a signature database after being processed by PCA. This technique will help strengthen the misuse IDS and increase its zero-day attack detection rate.[43,44]

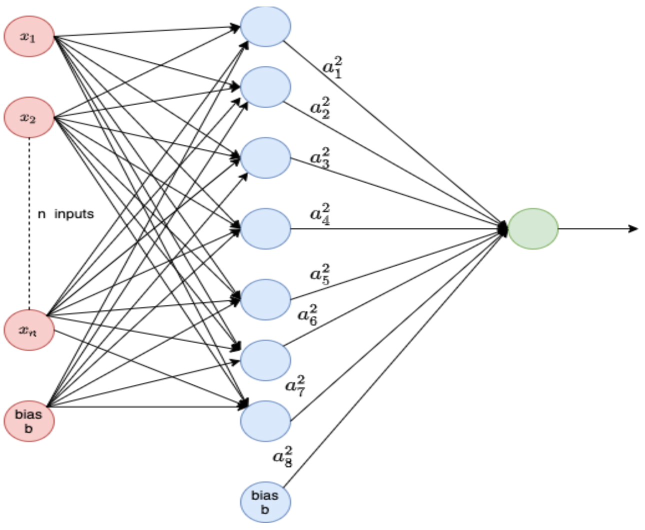

Figure 2 : Architecture of Neural Network

4) *Online Incremental SVM :*

To protect a network of IoT devices we will rely on an online Incremental SVM on a fog architecture installed on Raspberry pi. This misuse IDS will use a database that contains attack signatures coming from an intelligent honeypot and behavioural IDS processed by PCA to reduce data complexity for the SVM to perform better [40]. The SVM is a statistical supervised learning algorithm capable of doing binary classification, it uses a decision hyperplane line to maximize distance separation between two classes. For higher dimensional input data x_1,x_2,x_3,x_4……..x_n a kernel function is applied K and bias b with coefficient $\alpha_{0,i}$ as follows :

$$f(x) = \text{sign}(\sum_{i=1}^{n} \alpha_{0,i} K(x, x_i) + b) \qquad (3)$$

multiple kernel functions can be used like linear, RBF and polynomial kernel, but for this IDS we chose linear kernel function [45] :

$$K(x,x_i) = x^T * x_i \qquad (4)$$

To secure IoT internal network we will first use data PR_0captured by the honeypot, then we will apply PCA to reduce data dimensionality in order to create our initial model. This misuse IDS will permanently update its attack signatures by using an online database. To implement the online training we will use the initial vectors SV_0 calculated in first SVM model and add it to PR_1 to get another support vectors SV_1,we will follow this procedure recursively for every sample of data untill PR_n in signature database using SV_{n-1}vectors [45].The algorithm used in the online training is as follows :

$$SV_i = PR_i + SV_{i-1} \qquad (5)$$

IV. METRICS AND EVALUATION :

To establish a classification method we need an approach to measure the performance and relevance of a classification model.

To verify this system we need the following information:

- **True positive**: a positive represents a sample which was malicious and it has been well classified as malicious by the machine learning algorithm (30: see *"Table I"*).

- **True negative**: represents a sample which was correctly classified as Begnin (820: see *"Table I",*).

- **False positive**: are the data that was incorrectly classified as malicious even if they are Begnin (20: see *"Table I"*).

- **False negative**: represents data that was incorrectly classified as Benign while they are malicious. (30:see *"Table I"*).

- **Accuracy** : (TP + TN) / (TP + TN + FP + FN) = 94,44%.

In this example we found that the rate is high (94,44%), but does not reflect the quality of the model especially that among 60 malicious data 30 were well classified which represent 50%. The accuracy rate is a somewhat naive measure, it only gives a global vision of the model, but it is not particularly relevant for unbalanced data. This is why we will use the following metrics:

- **Precision**: TP / TP + FP = 30/50 (60%) it means when our model classifies a data as an attack and predicts it correctly with a rate of 60%.

- **Recall or TPR Sensitivity**: TP / TP + FN = 30/60 (50%) from all the data that was classified as malicious it represents the rate of what was definitely malicious.

- **False positive rate (FPR)**: FP / FP + TN = 20/840 (2,38%) is the rate of elements that have been misclassified among the normal data (True negative).

V. CONCLUSION AND IMPLEMENTATION CHALLENGES

IoT devices are known for their limited resources such as memory or processing time, which has forced us to use a fog computing architecture to protect the network from attacks, combined with cloud computing to leverage storage and processing power to perform complex tasks such as reducing large data features. The framework described in this article is adaptive to new attacks especially because it gets updated using live data captured from the internet, which gives it an edge advantage over other types of IDS that are trained either on data not intended for IoT devices or do not use data that gives better results for new attacks. We tried to combine high true positive of misuse IDS and increase detection rate by getting better data quality captured directly from the internet using an intelligent honeypot. We have to implement this framework and compare its performance with other type of IDS to make sure this misuse IDS can perform better in detecting zero day attacks in challenging environment.

ACKNOWLEDGMENT

The authors would like to thank our colleagues in mathematics and statistics laboratory at the university Mohamed V who contributed with their insights and mastery of the subject to assist this research

Table I. Confusion Matrix

	Classification(Malicious)	**Classification(Benign)**
Real(Malicious)	**30 (True positive)**	**30 (False negative)**
Real (Benign)	**20 (False positive)**	**820 (True negative)**

We would also like to show our gratitude to Shodan scanning website for giving us access to their API that helped in creating this paper.

REFERENCES

[1] M. Noura, M Atiquzaman, M Gaedke Interoperability in internet of things: Taxonomies and open challenges.Mobile Networks and Applications, 2019.

[2] E Al Nuaimi, H Al Neyadi, N Mohamed Applications of big data to smart cities. Journal of Internet,2015.

[3] J Gubbi, R Buyya, S Marusic, M Palaniswami Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 2013.

[4] Zhongjin Liu, Le Zhang, Qiuying Ni, Juntai Chen, Ru Wang, Ye Li, Yueying He An Integrated Architecture for IoT Malware Analysis and Detection. International Conference on Internet of Things as a Service, 2018.

[5] Mohsen Marjani, Fariza Nasaruddin, Abdullah Gani, Ahmad Karim,Ibrahum Abaker,Targio Hashem,Aisha Siddiqa,Ibrar Yaqoob Big IoT Data Analytics : Architecture, Opportunities, and Open Research Challenges, 2017.

[6] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic,Maritmuthu Palansiwami. Internet of things (IOT) : A Vision, Architectural Elements, and Future Directions. Future Generation Computer Systems, 2013.

[7] Summia Taj, Uniza Asad, Moeen Azhar,Sumaira Kausar. Interoperability in IOT based smart home : A review. Review of Computer Engineering StudiesVol.5, No.3, September, 2018, pp. 50-55.

[8] Luigi Atzori, Antonio Iera,Giacomo Morabito. Internet of things a survey. Computer Networks Volume 54, Issue 15, 28 October 2010, Pages 2787-2805.

[9] Anna L. Buczak,Erhan guven A survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications surveys & tutorials, 2015.

[10] MY Su Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers Expert Systems with Applications, 2011. 38(4) : p. 3492-3498

[11] https://resources.infosecinstitute.com/network-design-firewall-idsips/

[12] C Livadas, R Walsh, DE Lapsley, WT Strayer Using Machine Learning Techniques to Identify Botnet Traffic.Proceedings. 2006 31st IEEE Conference on Local Computer Networks.

[13] F Jemili, M Zaghdoud, MB Ahmed A Framework for an Adaptive Intrusion Detection System using Bayesian Network.Intelligence and Security Informatics, ISI, IEEE International Conference, 2007.

[14] C Kruegel, D Mutz, W Robertson Bayesian event classification for intrusion detection. 19th Annual Computer Security Applications Conference, 2003.

[15] S Benferhat, T Kenaza A naive bayes approach for detecting coordinated attacks. 2008 32nd Annual IEEE International Computer Software and Applications Conference.

[16] M Panda, MR Patra Network intrusion detection using naive bayes. IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.

[17] NB Amor, S Benferhat, Z Elouedi Naive bayes vs decision trees in intrusion detection systems, 2004.

[18] C Kruegel, T Toth Using decision trees to improve signature-based intrusion detection. International Workshop on Recent Advances in Intrusion Detection, 2003.

[19] http://manual-snort-org.s3-website-us-east-1.amazonaws.com/

[20] http://cedric.cnam.fr/vertigo/Cours/ml2/coursArbresDecision.html

[21] L Bilge, E Kirda, C Kruegel, M Balduzzi, EXPOSURE : Finding Malicious Domains Using Passive DNS Analysis, 2011.

[22] L Bilge, S Sen, D Balzarotti, E Kirda, Exposure: A passive dns analysis service to detect and report malicious domains. ACM Transactions on Information and System Security April 2014.

[23] GR Hendry, SJ Yang, Intrusion signature creation via clustering anomalies. Proceedings Volume 6973, Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2008.

[24] Tongbo Luo, Zhaoyan Xu, Towards an Intelligent-Interaction Honeypot for IoT Devices : IoTCandyJar. Black Hat, 2017.

[25] https://blog.avast.com/mqtt-vulnerabilities-hacking-smart-homes

[26] https://blog.shodan.io/security-researchers-find-vulnerable-IoT-devices-and-mongodb-databases-exposing-corporate-data/

[27] ML Bringer, CA Chelmecki, H Fujinoki A survey: Recent advances and future trends in honeypot research. I.J.Computer Network and Information Security, 2012,10, 63-75.

[28] Hazzaa Alshareef ; Marwah Almasri ; Abdulaziz Albesher ; Dan Grigoras Towards an Effective Management of IoT by Integrating Cloud and Fog Computing. IEEE International Conference on Smart Internet of Things (SmartIoT), 2019.

[29] D Nallaperuma, R Nawaratne, Online incremental machine learning platform for big data-driven smart traffic management. IEEE Transactions on Intelligent Transportation Systems (Volume: 20 , Issue: 12 , Dec. 2019).

[30] H.H Pajouh, R, Javidan, R, Khayami, D. Ali, and K. K. R. Choo, « A Two-layer Dimension Reduction and Two-tier Classification Model for Anomaly-Based Intrusion Detection in IoT backbone networks » IEEE Transactions on emerging Topics in Computing, vol. PP, no 99,pp. 1-1, Nov. 2016.

[31] D. Kushner, ''The real story of stuxnet,'' *IEEE Spectr.*, vol. 50, no. 3, pp. 48–53, Mar. 2013

[32] Elike Hodo, Xavier Bellekens, Andrew Hamilton, Pierre-Louis Dubouilh. Threat analysis of IoT networks using artificial neural network intrusion detection system. International Symposium on Networks, Computers and Communications (ISNCC), 2016

[33] Peiying Tao,Zhe Sun, and Zhixin Sun. An improved Intrusion Detection Algorithm based on ga and svm. IEEE Access, 6:13624-13631,2018.

[34] Shadi Aljawarneh,Monther Aldwairi, and Muneer Bani Yassein.Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 25:152-160,2018.

[35] Yang Li,Jun Li Wang,Zhi-Hong Tian,Tian-Bo Lu, and chen Young.Building lightweight intrusion detection system using wrapper-based feature selection mechanisms Computers & Security, 28(6): 466-475, 2009.

[36] SANA ULLAH JAN, SAEED AHMED, VLADIMIR SHAKHOV,INSOO KOO. Towards a Lightweight Intrusion Detection System for the Internet of Things.IEEE Access,7:42450 - 42471,2019.

[37] JESUS PACHECO, VICTOR H . BENITEZ , LUIS C . FÉLIX-HERRÁN, AND PRATIK SATAMArtificial Neural Networks-Based Intrusion Detection System for Internet of Things Fog Nodes. IEEE Access,8:73907 - 73918,2020.

[38] A. S. Sohal, R. Sandhu, S. K. Sood, and V. Chang. A cybersecurity framework to identify malicious edge device in fogcomputing and cloud- of-things environments.Computers & Security. 2019

[39] http://www.jmlr.org/papers/v7/laskov06a.html

[40] Cheng-Tao Chu,Sang Kyun Kim,Yi-An Lin,YuanYuan Yu,Gary Bradski,Andrew Y NG,Kunle Olukotun. Map-Reducefor Machine Learning on Multicore.Advances in Neural Information Processing Systems 19 (NIPS 2006)

[41] https://www.guru99.com/introduction-to-mapreduce.html

[42] Wei Wang, Roberto Battiti. Identifying Intrusions in Computer Networkswith Principal Component Analysis First International Conference on Availability, Reliability and Security (ARES'06), 2006.

[43] Elike Hodo, Xavier Bellekens, Andrew Hamilton, Pierre-Louis Dubouilh, Ephraim Iorkyase, Christos Tachtatzis,Robert Atkinson. Threat analysis of IoT networks Using Artificial Neural Network Intrusion Detection System.International Symposium on Networks, Computers and Communications (ISNCC),2016

[44] Sanmeet Kaur, Maninder Singh.Hybrid intrusion detection and signature generation using Deep Recurrent Neural Networks. Springer Neural Computing and Applications, 2019.

[45] Nadeem Ahmed Sayed,Huan Liu,Kah Kay sung.Incremental Learning with Support vector Machines, 1999.

[46] Xu, Q., Aung, K. M. M., Zhu, Y., & Yong, K. L.

A Blockchain-Based Storage System for Data Analytics in the Internet of Things. Studies in Computational Intelligence, (Springer) 119–138, 2017.

[47] Cristiano Antonio de Souzaa,Carlos Becker Westphall,Renato Bobsin Machado,João Bosco Mangueira Sobral,Gustavo dos Santos Vieira. Hybrid approach to intrusion detection in fog-based IoT environments, Computer Networks,Elservier,2020.

ABDELLATIF EL GHAZI
He is currently a professor since 2012 at school of energy of UIR, and member of TICLab. he coordinates two projects, ERASMUS+ e-VAL and MarMooc.
His research interests include :
digital analysis and optimization, cloud computing, IT security, IoT and artificial intelligence.

RACHID AIT MOULAY
Received his Master degree in software engineering from the University Mohamed V of Rabat in 2010.
He served about 10 years as a software Engineer engineer in many companies.
He is currently pursuing Ph.D from the University Mohamed V of Rabat.
His research areas include :
Cyber Security,IoT,Machine Learning,Smart Homes.