Machine Learning Techniques for Traffic Identification and Classification in SDWSN: A *Survey*

Ratanang Thupae
Department of Computer Science
North-West University
Mafikeng, South Africa
ratanangthupael@gmail.com

Bassey Isong
Department of Computer Science
North-West University
Mafikeng, South Africa
isong.bassey@ieee.org

Naison Gasela
Department of Computer Science
North-West University
Mafikeng, South Africa
naison.gasela@nwu.ac.za

Adnan M. Abu-Mahfouz
Modelling and Digital Science
CSIR
Pretoria, South Africa
a.abumahfouz@ieee.org

***Abstract*—Software defined network (SDN) is a paradigm developed achieve great flexibility and cope with the limitations of traditional networks architecture such as the wireless sensor networks (WSNs). Introducing SDN in WSN leads to SDWSN. However, due to the challenges that are inherent in SDN and WSN, SDWSN is faced with number of challenges such network and Internet traffic classification (TC). Several solutions have been offered such as machine learning (ML) technique but there are several challenges that still exist which need attention. Therefore, this paper present a review on the approaches of TC in SDWSN using ML and their challenges. The objective is to identify existing approaches and the challenges in order to provide ways to enhance them. We performed review of the existing works on TC in the literature based on the aspect of enterprises network, SDN and WSN has been done as well as findings reported. Our findings shows that the approaches to TC using ML were based on supervised or unsupervised learning. Moreover, TC is faced with challenges which include energy efficiency, shareable test data and design. Thus, ML technique to TC in SDWSN is still at its early stage and need to improve in order to accurately classify traffics that normal or abnormal.**

Keywords— WSN, SDWSN, Traffic, Classification, Machine learning, Security.

I. INTRODUCTION

Software defined network (SDN) is an emerging technology in the realm of computer networks and areas of telecommunication. Its goal is to solve challenges in Internet protocol (IP) based networks[1] and is based on the idea that decouples data plane from the control plane. In particular, the responsibility of data plane is to forward data packets and the control plane ensures efficient data routing, traffic engineering (TE) and policies management. OpenFlow is one of the significant protocols which allows the controller to interact with network switches and it is a standard interface that is utilized in SDN. One of the application of SDN is in wireless sensor network (WSN) to enhance its network management and control. This association yields software defined wireless sensor network (SDWSN). Despite the benefits that SDWSN, the paradigm is threaten by challenges that were mostly inherited from SDN and WSN.

In the SDWSN, to ensure high availability and efficiency in network management, traffic classification (TC) is essential. Two common approaches are usage of port numbers (PNs) and deep packets inspection (DPI) are important when classifying traffics in networks [2]. In previous works, these methods were considered ineffective considering the nature of modern networks and quickly deviated their focus to machine leaning (ML) techniques which is based on statistical properties. Although there are still some challenges on TC based on ML, SDN play a critical role in network management in terms of global view from controllers [3, 4]. The extraction of traffic based on statistical data from switches is easy to use because it can be done in an offline or online manner. However, there two identified challenges originating from the online approach such as high time complexity and processing overhead [1]. In the perspective of WSN and SDN, network detection technique constitute the first step to identify and classify type network classes. This play a fundamental role in network management and security, intrusion and quality of service (QoS) and so on [5].

There are several techniques for TC that exist. For instance, port based, payload based were proposed by Nguyen *et al.* and Foremski *et al.* [6, 7] in order to classify unknown classes. Shafiq *et al.* [5] stated that the valid reason for introducing ML techniques is due to the fact that the port based technique failed due to peer to peer (P2P) applications which normally utilize dynamic port numbers. Moreover, payload based approach generates accurate results in terms of network TC but cannot be used for encrypted data network applications and this technique experienced failure due to utilization of encrypted flow of applications. Due to these drawbacks, ML techniques were proposed by researchers to classify traffics flows and also to identify types of applications in the network.

Due to its promising accuracy, SDN in WSN is geared towards addressing the challenges of TC [8]. However, TC techniques utilize concept of flow which are defined as group of packets with the same Internet protocol (IP) addresses, transport protocol (TP) and its port numbers [9]. In WSN, TC is still challenging because network traffic increase load of work due under performing personnel who cannot work properly. Therefore, this paper bring together some of the potential issues of network traffic identification and classification in SDWSN and countermeasures already proposed. We surveyed and analysed some of the existing works and presented summary of the identified challenges.

TABLE I. SUMMARY OF MACHINE LEARNING BENEFITS IN SDWSN

Ref.	Machine Learning Algorithm(s)	Software (Tools used)	Traffic classes considered
Shafiq *et al* [5]	C4.5, Support vector machine (SVM) , Bayes Network and naïve Bayes	Wireshark for capturing network traffic , NetMate for extraction of features and Weka [24] for classifying network traffic **Alternative:** Tcpdump for capturing real time network traffic, Perl script for extraction of features from captured data set and MatLab for classifying network traffic	WWW, DNS, FTP, P2P and Tenet applications
Sadawarte *et al* [25]	Naïve Bayes	-	Database, mail service, P2P, Bulk data transfer, Multimedia and Games
Fruhwirt *et al* [19]	-	Wireshark [19] for capturing network traffic, Weka [24] for traffic classification using JBBC- database **Alternative** Parsing script written in Lua	HTTP, etc
Liu Zhen and Liu Qiong [11]	Naïve Bayes	Weka for traffic classification	-
Murat Soysal and Ece Guran Schmidt [26]	Bayesian Networks, Decision Trees and Multilayer perceptions	Weka for traffic classification	P2P, Web(HTTP), Akamai, FTP, DNS and SMTP

The implication is that future researcher will be guided accordingly in providing efficient solutions to network TC.

The remaining parts of the paper is organized as follows: Sect. II presents related works, Sect. III presents the approaches and issues of traffic classification in SDWSN, Sect. IV is discussion of the paper and Sect. V is the conclusion.

II. RELATED WORKS

This section presents some of the related works that have been performed on detection and classification of network traffic in SDWSN. They are discussed as follows:

Namdev *et al.* [10] and Zhen *et al.* [11] stated that ML does not only solve the concerning issues of traditional techniques but it improves efficiency regarding application of supervised and unsupervised ML techniques. The study also pointed out some improvements in terms of processing speed, accuracy that need attention via increasing the size of dataset. In addition, they reported that TC relied on transport control protocol (TCP) and user datagram protocol (UDP) for a long time. ML techniques learn from big data and utilize statistical properties of the traffic flow. To ensure efficient TC, there are many principles that need to be followed and implemented. Thus, for accurate and efficient TC, efficient architectural models need to be introduced by research community. Wang *et al.* [12] presented issues concerning TC based on quality of service (QoS) which concentrated on unique architectural features of SDN. Their findings shows unsupervised ML algorithms cannot be directly applied in SDN. However, they suggested that for application, separated control and data planes need to exploited and conformed.

Namdev *et al.* [10] also presented their work on traffic classification (TC) utilizing ML techniques. They found out that ML techniques don't only overcome traditional techniques problems but it improves its efficiency. They also stated that although many supervised and unsupervised ML techniques have been applied and also need a direction to improve the accuracy. Jamuna and Vinoth [13] presented their work in the field of ML focusing on TC with a reason to move away from payload or port based TC. Their findings show ML is more efficient in the task of classifying traffic. The study suggested that Autoclass, Expectation maximization, Decision Tree and Naïve Bayes ML algorithms can be used for offline analysis due to demonstrated high accuracy within the Internet application traffic. In a similar study, Zhou *et al.* [16] explored new ways for classifying network traffic based on ML techniques. They presented new approach to classify traffic regarding statistical properties (i.e. five-dimensional tuple unidirectional and bidirectional flows) utilizing duration of the flow, total length of the flow, number of packets sent and inter-packet arrival times etc. They concluded bidirectional flows was more active unlike unidirectional flows in TC.

III. SECURITY ISSUES IN SDN-SDWSN

Traffic classification is the process of identifying the applications or protocols in a network. This is a significant technique that has been used for more than two decades and many methods for classifying network applications have been proposed. This section presents an overview and issues of network and Internet TC using artificial intelligence (AI) approach also known as ML in SDWSN. They are as follows:

A. Machine Learning

ML is a powerful technique categorized into the following learning divisions namely the supervised learning, unsupervised learning and semi-supervised learning (hybrid or combination of techniques). There is also a new Internet network TC technique known as genetic algorithm (GA). In contrast, GA techniques [14] have application areas in ML

approaches from AI system such as crossing, mutation and selection stages as operations utilized in solving problem that arises. Moreover, identification of feature vector, estimation of statistical distribution of features vector and defining the example are the three ML methods to be considered when classifying network or internet traffic [6, 15].

1) Supervised Learning techniques: This is a classification method where data set is completely labelled to classify unknown classes. It operates by training a model with some of labelled data set to produce predicted output in new data samples [5]. However, in this approach, classification is done using two steps, training and testing and it examine data provided and classification model built in the training step classifies new unseen instances [10]. Some of the algorithms in this category are: Namdev *et al. and* Zhou *et al.* [10] [16] presented these algorithms as follows: Neural network (Feed forward), an approach proposed for accuracy of TC in combination with fast correlation based feature and feature selection algorithm was utilized for eliminating redundant features. However, this approach work as a classifier. Neural network (Radial Basis Function-RBF) is regarded as multi-layer feed forward artificial neural network (ANN) which utilizes radial basis functions and is categorized into three layers architecture such as hidden layer with a non-linear RBF function etc. Bayes Network is known as a probabilistic model using the graph model to represent set of random variables and it also use the directed acyclic graph (DAG) to represent those sets. It consists of two steps, first one is a learning of network structure and second is estimate probabilistic table based on random variable. Naïve Bayes trees is classified as a technique based on applying bayesian theorem with strong and independence assumption and it assumes presence or absence of a certain feature of a class. C4.5 Decision Tree is a well-known decision-based ML algorithm utilized to develop univariate decision tree. Also an enhancement of iterative dichotomiser 3 (ID3) algorithm is utilized to find simple decision tree. It is also a statistical classifier due to its good ability in classification. Support vector machine (SVM) [17] is another approach used for statistical learning theory (STL) where the classification of the low dimensional space can be transferred into higher dimensional one. In SVM performance based on computation, the best solution is designed by selecting appropriate kernel functions [13].

2) Unsupervised Learning techniques: This technique is called a cluster because there is no need for data sets to be labelled and result does not identify instances in predefined classes. ML approaches take a lesson from big data perspective and utilize statistical properties (SP) of the traffic flows to infer the application utilized. However, Dainotti *et al.* [18] stated that algorithms based on pattern recognition field utilizing ML techniques have given promising solutions especially in encrypted traffic which involve payload analysis. ML is based on AI technique which offers flexibility on this reaction to new situations and also retrain to fight intrusions or attacks [19]. The ML has great potential and theory in solving many Internet service providers (ISPs) problems. Matlou *et al.* [20] state that ML from AI perspective is still used by many scientists and engineers whom specify it as a significant technique in WSN due to their dynamic environment and data collected. Namdev *et al.* [10] presented the this algorithms as follows: *K-Means* is a clustering algorithm which partition dataset objects into K disjoined subsets and maximizes the homogeneity of the cluster and is also in control of maximizing the square-error based on the distance between each object and the mean of the object. Auto-class is a proposed algorithm known as iterative method used to maximize the parameters in order to produce clusters and it consist of two steps namely expectation and maximization. The first step estimate parameter using random numbers and secondly uses variance by re-estimating that parameter. Also is a density-based spatial clustering of application with noise (DBSCAN) is classified as a data clustering algorithm and also a density based clustering algorithm which finds the number of clusters initially from estimated density distribution of the corresponding nodes. Expectation maximization is a simple, practical and iterative algorithm which does not directly maximize or simulate complex posterior distribution. For its computation to be simplified some potential data based on observing data need to be included and executes a series of simulation [13].

B. Traffic Classification

Mirsky *et al.* [12] presented the framework deploying network intrusion detection system (NIDS) deployed at the internet gateway (single point). This strategy is useful in terms of detecting malicious traffic entering and leaving the network. Nonetheless, the concerning issue is that those malicious attacks cannot travel across the network itself. They suggested that an implementation strategy based on distributed point where number of NIDs will be connected to routers and gateways within the network. They used artificial neural network (ANN) which have an advantage in the detection performance using other ML algorithms. Arsalan *et al.* and Zhang *et al.* [13, 14] presented a framework determining existing flows of application type in wireless network. Based on their findings, when host joined the network, device flow properties in terms of TC model were sent by OpenFlow (OF), however, in terms of ML technique, C5.0 decision tree algorithm was the one used and showed promising accuracy. Moreover, the study stated several challenges in recent networks based on TC and global view of controllers in SDN which can improve the network management. They suggested a method using the flow properties gathered in a dataset such as K-means algorithm input. Fruhwirt *et al.* [15] presented several ML techniques to possibly classify new traffic that traverse the network. They state that the usage of these techniques have the capability of offering many Meta information. In predicting, the next steps of the attacker that need to be considered are important because to determine a host their approach doesn't necessarily utilize Internet protocol (IP) addresses. Hence, they suggested that existing approach to cluster traffic or splitting it up based on different attributes need to be extended because using different attribute than IP address would possible to count those hosts behind network address resolution (NAR).

IV. Mechanisms and Issues in Traffic Classification

This section presents the fundamental mechanisms and issues experienced in TC.

A. Mechanisms

Existing approaches are discussed as follows:

1) Intrusion detection system: This is a system utilized to monitor network traffic by looking for malicious attacks or activities [21, 22]. Fruhwirt *et al.* [19] stated that this mechanism is based on assumption, there is a different behaviour coming from intruders compared to an unsuspicious user. They further elaborated that challenge for IDS mainly differs from two types of behaviour where intruder's behaviour shows false positives and normal user traffic as false negatives.

2) Network based defence: It is clear that security still pose as an issue and need possible attention because complete unsecured system can be vulnerable to attacks. This mechanism is a very significant countermeasure that add additional security layer to the network which somehow minimizes the risk of attacks [19].

B. Issues

Some of the issues that impacts TC negation are as follows:

1) Energy efficiency: Wang et al. [12] reviewed energy efficiency issue under SDN perspective and they stated that if given proper attention, it will bring capability in terms of high performance. They further explained that methods need to be implemented in SDN and WSN to accommodate SDWSN realm. In particular, this is for controlling inflating energy consumption while proving security to the entire network. They also emphasized that implementing energy consumption strategies in SDN could reduce overall energy usage and this could be beneficial to TC or management. Moreover, they stated that sleep-awake mechanism called SDN-ECCKN in SDN could be implemented due to its capability of reducing power consumption.

2) Shareable test data: Privacy is one of the constraints when data is disclosed, makes it hard for researchers in attempting to improve the way traffic is classified. Vladutu et al. [23] stated that the concerning issue is shareable test data input which is aligned with flow objects that are re-labelled to be utilized as reference. The study also found important capacity of scalable buildings, parallel traffic classifiers and they will be able to deal with huge powerful data traversing the context of networks.

3) Design and requirement: Deep packet inspection (DPI) is classical technique used to classify network traffic through IDS. This device should be periodically updated with new traffic patterns. However, this operation is significant in terms of malicious traffic like viruses. This can be challenging issue because illegitimate traffic can take control of the new masks overnight and the traffic pattern update need to occur quickly. This can be difficult for a network administrator due to the tedious job because in this situation many errors can occur.

V. Discussions

In this paper, we have highlighted some of the existing concerning issues faced by network and Internet TC. This paper found that SDWSN is faced with several problems in traffic identification and classification. Moreover, several approaches exist and each has its own strengths and weaknesses. However, the trending technique is the ML-based approach in which there are several algorithms existing. The summary of the ML algorithms technique is shown in Table I and Table II while Table III captures the existing challenges in the aspect of Traffic identification and classification.

ML algorithms such as C4.5, SVM, Bayes Network and Naïve Bayes which haven applied so far produce promising accuracy in terms of performance. In addition to the tools used, Wire shack was used to capture traffic, NetMate for extraction of features and Weka for classifying traffic on the network based on various traffic classes (See Table I and II). In terms of the challenges, energy efficiency has been considered one of significant challenges found in TC and has not yet been addressed in its architectural design. However, if this issue cannot be given enough attention by researchers, there is a possibility that SDWSN system can suffer from fault tolerance, scalability and reliability issues. Moreover, there has always been an argument that with introduction of SDN can help solve issues faced by WSN. ML is one of the recent solutions that became popular with various research communities for classifying network and Internet traffic, particularly for detecting any traffic anomalies. TC is important in the deployment of several network activities such as traffic engineering (TE), quality of service (QoS) and anomaly behaviour detection [6]. In contrast, detection of traffic changes is a significant operation to adjust the resources accordingly with additional costs. Basic functions provided to Internet service providers (ISPs) for example can be utilized for IDS by finding patterns of DoS attack and so on.

Researches over the years have shown that ML can be very helpful in terms of TC. The two points which are traffic classification and detection of network changes can be done using statistical properties. In particular, analysing statistical properties is utilized by different routers that detect networks. However, techniques like deep packet inspection (DPI) has been used for a long time but constitute time consuming as compared to the ML approaches. ML techniques can automatically determine traffic pattern changes which are exploited by network controller detection. Moreover, despite the edge ML approaches has, it is also marred with several challenges as shown on Table III. In a nutshell, the ML algorithms brings new approach to TC giving them capability to be sustainable when faced with issues. The network detection and classification techniques using ML algorithms when applied to SDWSN will improve TC accuracy. However, the possible challenges when applying technique(s) to SDWSN is the occurrence of new application if traffic can be well classified and due to energy consumption network can experience efficient data flows. Therefore, choosing inaccurate or inconsistent ML algorithms may result to a very challenging task regarding TC.

TABLE II. MACHINE LEARNING ALGORITHM PERFORMANCE

Ref.	Machine Learning Technique	Classification method	Feature selection algorithms	Accuracy	Dataset used
Zhang *et al*, Kaur *et al* Singh *et al* [27-29]	Supervised	SVM, C4.5 Decision Tree, Naïve Bayes, Bayes network, Naïve Bayes Tree, Random Forest and C4.5	Principal component analysis, Consistency and Correlation	90 % - 94%	Proprietary hand classified traces, UTM campus network, Educational institution and The Auckland
Hamza *et al* , Shi *et al* [30, 31]	Unsupervised	DBSCAN, K-Means	Fast correlation based filter, Principal component analysis, Consistency and Correlation	90% - 97 %	The Auckland, Proprietary hand classified traces and UTM campus.

TABLE III. MACHINE LEARING-BASED TRAFFIC CLASSIFICATION SUMMARY

Ref.	Application	Technique used	Challenge	Objectives
Wang *et al* [12]	Traffic classification (Network)	Machine Learning (Laplacian support vector machine algorithm from supervised learning)	Based on new connected devices it is difficult to keep track of all data.	For energy consumption data flows need proper classification.
Mirsky *et al* [32]	Traffic identification and classification (Internet and network) using network intrusion detection (NID) system	Artificial neural network-KitNET algorithm using Kitsune	Based on single point deployment, detection of malicious traffic entering and leaving the network is a problem but not malicious traffic that travelling across the network.	However, for the problem to be solved, distributed deployment is needed.
Liu Zhen and Liu Qiong [11]	Traffic classification (internet)	Machine Learning (Supervised discretize method in WEKA).	Measurement of bias degree of one feature in one class.	This challenged can be addressed by evaluating bias coefficient in terms of information theory.
Vl˘adu¸tu *et al* [23]	Traffic classification (Internet)	-	Shareable test data input which is a concerning issue	-

VI. CONCLUSION

SDWSN is still an emerging paradigm, with its support from traditional network such as SDN and WSN. However, this technology is faced with many challenges. SDWSN is experiencing some concerning issues of traffic identification and classification. Thus, the network need to be efficiently managed, monitored and secured. This paper presented some of the challenges inherited from SDN. Particularly in the literature. We surveyed those challenges and some of the proposed objectives on the aspect of energy efficiency, shareable test data, design and so on. They constitute some of the concerns which need researcher‘s attention both in industry and academia. Our findings shows that the application of ML techniques in the identification and classification of network traffic in SDWSN is still at its developmental stage. Each algorithms used has its merits and demerits which have implication to their choice of usage. The future work is to apply ML techniques in SDWSN using several ML algorithms to determine which technique is best due to their performance. Also, such ML will be used to design an efficient TC framework, implemented and validated based on several issues identified in this paper.

ACKNOWLEDGMENT

This research was supported by FRC and the Department of Computer Science at the NWU-Mafikeng and CSIR, South Africa.

REFERENCES

[1] Parsaei, M.R., et al., Network Traffic Classification using Machine Learning Techniques over Software Defined Networks. International Journal Of Advanced Computer Science And Applications, 2017. 8(7): P. 220-225.

[2] Parvat, T.J. and P. Chandra, A Novel approach to deep packet inspection for intrusion detection. Procedia Computer Science, 2015. 45: p. 506-513.

[3] Williams, N., S. Zander, and G. Armitage, A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. ACM SIGCOMM Computer Communication Review, 2006. 36(5): p. 5-16.

[4] Kobo, H.I., A.M. Abu-Mahfouz, and G.P. Hancke, Fragmentation-based Distributed Control System for Software Defined Wireless Sensor Networks. IEEE Transactions on Industrial Informatics, 2018.

[5] Shafiq, M., et al. Network traffic classification techniques and comparative analysis using machine learning algorithms. in Computer and Communications (ICCC), 2016 2nd IEEE International Conference on. 2016. IEEE.

[6] Nguyen, T.T. and G. Armitage, A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys & Tutorials, 2008. 10(4): p. 56-76.

[7] Foremski, P., On different ways to classify Internet traffic: a short review of selected publications. Theoretical and Applied Informatics, 2013. 25.

[8] Akpakwu, G.A., et al., A survey on 5G networks for the Internet of Things: Communication technologies and challenges. IEEE Access, 2018. 6: p. 3619-3647.

[9] Cai, J., Z. Zhang, and X. Song. An analysis of UDP traffic classification. in Communication Technology (ICCT), 2010 12th IEEE International Conference on. 2010. IEEE.

[10] Namdev, N., S. Agrawal, and S. Silkari, Recent advancement in machine learning based internet traffic classification. Procedia Computer Science, 2015. 60: p. 784-791.

[11] Zhen, L. and L. Qiong, A new feature selection method for internet traffic classification using ml. Physics Procedia, 2012. 33: p. 1338-1345.

[12] Wang, P., S.-C. Lin, and M. Luo. A framework for QoS-aware traffic classification using semi-supervised machine learning in SDNs. in Services Computing (SCC), 2016 IEEE International Conference on. 2016. IEEE.

[13] Jamuna, A. and V. Ewards, Survey of Traffic Classification using Machine Learning. International journal of advanced research in computer science, 2013. 4(4).

[14] Ertam, F. and E. Avcı, A new approach for internet traffic classification: GA-WK-ELM. Measurement, 2017. 95: p. 135-142.

[15] Yin, C., S. Li, and Q. Li, Network traffic classification via HMM under the guidance of syntactic structure. Computer Networks, 2012. 56(6): p. 1814-1825.

[16] Zhou, W., et al. Internet traffic classification using feed-forward neural network. in Computational Problem-Solving (ICCP), 2011 International Conference on. 2011. IEEE.

[17] Jing, N., et al. An efficient SVM-based method for multi-class network traffic classification. in Performance Computing and Communications Conference (IPCCC), 2011 IEEE 30th International. 2011. IEEE.

[18] Dainotti, A., A. Pescape, and K.C. Claffy, Issues and future directions in traffic classification. IEEE network, 2012. 26(1).

[19] Fruhwirt, P., S. Schrittwieser, and E. Weippl, Using machine learning techniques for tra c classi cation and preliminary surveying of an attacker's pro le. 2015.

[20] Matlou, O.G. and A.M. Abu-Mahfouz. Utilising artificial intelligence in software defined wireless sensor network. in Industrial Electronics Society, IECON 2017-43rd Annual Conference of the IEEE. 2017. IEEE.

[21] Kgogo, T., B. Isong, and A.M. Abu-Mahfouz. Software defined wireless sensor networks security challenges. in AFRICON, 2017 IEEE. 2017. IEEE.

[22] Pritchard, S.W., G.P. Hancke, and A.M. Abu-Mahfouz. Security in Software-Defined Wireless Sensor Networks: Threats, Challenges and Potential Solutions. in IEEE Int. Conf. of Ind. Informat., Emden, Germany. 2017.

[23] Vlăduţu, A., D. Comăneci, and C. Dobre, Internet traffic classification based on flows' statistical properties with machine learning. International Journal of Network Management, 2017. 27(3).

[24] Weka, W., 3: data mining software in Java. University of Waikato, Hamilton, New Zealand (www. cs. waikato. ac. nz/ml/weka), 2011. 19: p. 52.

[25] Sadawarte, D.S. and S.Y. Gaikwad, Survey On Network Traffic Classification Techniques With Correlation Information.

[26] Soysal, M. and E.G. Schmidt, Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison. Performance Evaluation, 2010. 67(6): p. 451-467.

[27] Zhang, J., et al., Internet traffic classification by aggregating correlated naive bayes predictions. IEEE Transactions on Information Forensics and Security, 2013. 8(1): p. 5-15.

[28] Kaur, J., S. Agrawal, and B. Sohi, Internet Traffic Classification for Educational Institutions Using Machine Learning. International Journal of Intelligent Systems and Applications, 2012. 4(8): p. 37.

[29] Singh, K. and S. Agrawal, Performance evaluation of five machine learning algorithms and three feature selection algorithms for ip traffic classification. IJCA Special Issue on Evolution in Networks and Computer Communications (1), 2011: p. 25-32.

[30] Ibrahim, H.A.H., et al., Taxonomy of machine learning algorithms to classify real time interactive applications. International Journal of Computer Networks and Wireless Communications, 2012. 2(1): p. 2012.

[31] Dong, S., D. Zhou, and W. Ding. The study of network traffic identification based on machine learning algorithm. in Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on. 2012. IEEE.

[32] Mirsky, Y., et al., Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. arXiv preprint arXiv:1802.09089, 2018.