Network Security Using Machine Learning

Onyecherelam henry agbaeze, Aditya Gaonkar and Akshay Dubey

***Abstract*— Securing a computer network has been a major challenge for IT professionals over the years. With more advances in technology and the rising dependency of companies on their network infrastructure, providing a secure network has never been more import. Intrusion detection systems have been extensively used but rule base systems are unable to deal with new attack vectors that have never been seen in the past and they have the tendency to flag too many packets as being malicious leaving the IT personnel with too much filter work to do. Machine learning and Artificial intelligence has proved to be one of the best tools for solving many decision or predictive problems. The idea that machines can learn from previous data has opened endless possibilities. It is this capability that is now being harnessed to solve the problem of security in networked systems. This paper takes a survey of some of the popular machine learning algorithms that has been implemented in practice to provide a much more secured network infrastructure and some of the strengths and likely weakness of each.**

I. Keywords

Machine learning, Decision Tree, Classification, Data log, Support Vector Machine, Hidden Markov Model.

II. INTRODUCTION

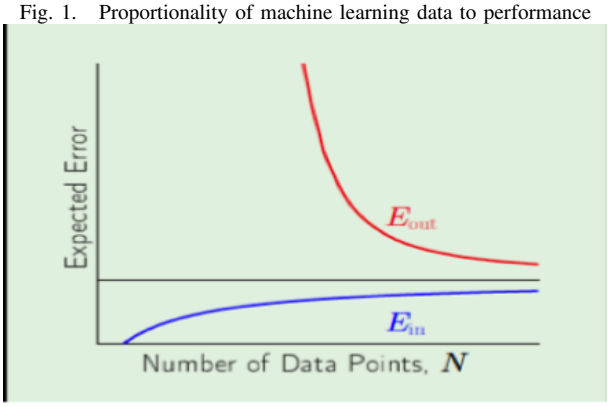
The need for securing a computer network has never been more important. Many companies rely heavily on network applications and also have to store sensitive data in the cloud.This means that securing the company network and data becomes more important than anything to both companies and individuals. A lot of research is has been done and is currently on going in using Machine learning algorithm techniques to tackle the issue of security.Classification of network traffic and building models that best represent a network's likely state at any point in time, is a problem that researchers try to solve. Many Machine learning algorithms have the characteristics to help with both detecting anomaly and misuse cases as well as detecting an attack signature more quickly. Another issue that has also been considered is that of accuracy. A single false positive or false negative can mean spending a lot of money, time and technical resources hence the best algorithms with the least error possible has to be implemented.

Using Neural network for real time security:

III. Data Sources for training ML models

There are a few ways we could access data which can be used to train machine learning models. This can be web server logs, windows event logs, Linux syslog, network logs, cloud service logs etc.[23]. According to [24] the more data we use to train the machine learning model, the better our prediction score. As we see from the fig 1 below, the more data points(N) we have, our solution model (Em) tends to get better while staying below the expected error.[24]

Fig. 1. Proportionality of machine learning data to performance



Using Neural network for real time security:

Neural network has been applied extensively in network classification. TCP payloads are made up of a sequence of hexadecimal numbers. Before training data is fed

into the neural network, the hex values are first represented as numbers between 0 -255 [23]. These values which are generated in real time by different network logs are then fed into the neural network as xi values in other to classify the network traffic accordingly depending on what features can be extracted by the neural network. The neural network keeps adjusting the values in the hidden layer until the weight on the hidden layer is minimise. This can be converted to an optimisation problem as seen in the equation below which is a form of Hebbian learning. The output is then refer into the neural network after IT personal have verified the correct classification of the network. This helps in training future model or leads to better classification [23].

Fig. 2. Hebbian equation

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

IV. INTRUSION DETECTION

Both Intrusion detection systems(IDS) and Intrusion prevention systems(IPS) have always been fundamentally part of the network security infrastructure for any computer networks.Certainly, implementation can take different forms but the main idea is usually the same. An intrusion detection system can easily be defined as a network technology which can be implemented in hardware or software for listening or monitoring network traffic and may take action if it detects any malicious activities or malicious traffic. This could be by sending an alert to the network administrator or taking initial measures to stop the attack.The network traffic usually may not pass through the IDS device but may listen to a network interface port. Intrusion prevention systems, on the other hand are built to detect and prevent a network vulnerability and prevent the attack before it happens.IPS examines the actual data packets that are in the network traffic and then tries to make intelligent predictions of a possibility of a malicious activity. This system are categorised in two parts: Anomaly-based detection and misuse-based detection

V. INTRUSTION DETECTION METHODS

Until recently, rule based anomaly detection has been used in computer networks to determine or detect any anomaly in the computer network. Network administrators would have to configure, specifically, what network IP or port numbers to permit or deny in a running network. This approach did work but still has a lot of limitations since an attack has to first occur before the source address can be flagged as malicious.Network design and architecture are always changing over time. Having to use single and direct rules which are generally not scalable lead to the system having lots of false positives and false negatives.This error margin is not acceptable since one since attack may mean millions if not Billions of money. Machine learning algorithms are now being used to detect and mitigate network level attacks on a network infrastructure. Even though this is a relatively new research are, a lot of progress have been made in using Artificial intelligence or Machine learning technique to solve network security challenges.

Until recently, rule based anomaly detection has been used in computer networks to determine or detect any anomaly in the

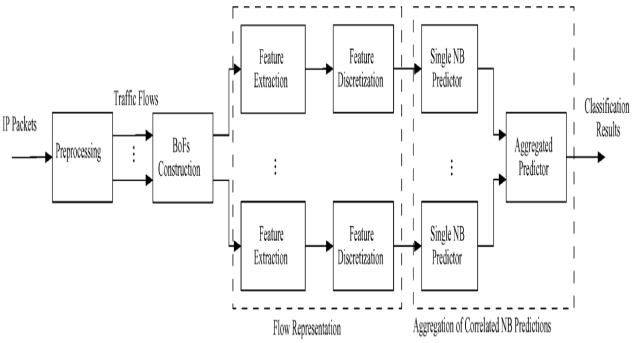
computer network. Network administrators would have to configure, specifically, what network IP or port numbers to permit or deny in a running network. This approach did work but still has a lot of limitations since an attack has to first occur before the source address can be flagged as malicious.Network design and architecture are always changing over time. Having to use single and direct rules which are generally not scalable lead to the system having lots of false positives and false negatives.This error margin is not acceptable since one since attack may mean millions if not Billions of money. Machine learning algorithms are now being used to detect and mitigate network level attacks on a network infrastructure. Even though this is a relatively new research are, a lot of progress have been made in using Artificial intelligence or Machine learning technique to solve network security challenges.

VI. TRAFFIC CLASSIFICATION BY AGGREGATING CORRELATED NAIVE BAYES PREDICTION

This approach deals with the correlated flows in an effective way, which shows significant performance even with a small set of supervised training data.

1) Process of classification: Fig. 3 explains the classification process of our proposed scheme, which is focused on flow-level traffic classification. In this process we analyse the IP packet headers across the desired target network and model our traffic flow. Specifically checking for 5- tuple, source IP, source port, destination IP, destination port, and transport layer protocol.We apply a heuristic way to determine the correlated flows and model them using "bag of flow"(BoF) concept. *"If the flows observed in a certain period of time share the same destination IP, destination port, and transport layer protocol, they are determined as correlated flows and form a BoF."*[1]. Classification is done by using a set of statistical features.These features are extracted and discretised to represent traffic flows. In this survey paper we discuss about classification approach named aggregation of correlated NB predictions[1]. This approach consists if two steps.*"In the first step, the single NB predictor produces the posteriori class-conditional probabilities for each flow. In the second step, the aggregated predictor aggregates the flow predictions to determine the final class for BoFs."*[1]

Fig. 3. classification process



2) BoF-Based Classification Framework: To model bag of flow set of correlated flows are generated by same application, X = $\{X_1....,X_b\}$. Since this flow belongs to same application it can help to improve the classification results. Research done in this paper[1] shows that goal can be achieved by following the approach of classifier combination.

Consider bag of flow classification problem X and m is number of possible traffic classes (w_1,....,w_m). Priori probability of oc-

currence is denoted by $P(w_k)$. Each class w_k modelled by the probability density function $p(X / w_k)$.

3) Aggregation of Correlated NB Predictions: In BoF-based NB approach, we aggregate correlated NB predictions[1], which results in a more accurate aggregated predictor for traffic classification.

4) Single NB Predictor: Naive Bayes classifiers provides two main advantages. Firstly, it has exhibits high classification speed and good performance using the discretised statistical features in traffic classification. Next, it is easy for naive Bayes classifier to produce the posterior probability that a testing flow belongs to a traffic class.

Bayesian decision theory [2] says that *"The maximum posterior classifier can minimise the average classification error. The key point is to estimate the posterior probability that a testing flow belongs to a traffic class"*. For any flow $x = \{x_1,...,x_n\}$, the posterior probability corresponding to class w is given in fig.

Fig. 4. Posterior Probability

$$P(\omega \mid \mathbf{x}) = P(\omega \mid x_1, \ldots, x_n).$$

Using Bayes theorem, we have

Fig. 5. Posterior Probability

$$P(\omega \mid x_1, \ldots, x_n) = \frac{P(\omega)p(x_1, \ldots, x_n \mid \omega)}{p(x_1, \ldots, x_n)}.$$

5) Single NB Predictor: According to Kittler's theoretical frame- work [3],*"A number of combination methods can be derived from the Bayesian decision theory which can be used for aggre- gated predictor."* Further mathematical proof and analysis can be found in this work[1].

6) Comparison With State-of-the-Art Methods: According to experiments conducted [1], performance comparisons were made between BoF-NB and other three methods: C4.5, k-NN, and Erman's semi supervised method [4]. C4.5 and k-NN demonstrate superior traffic classification performance in recent research [5], [6]. Erman's semi supervised method [4] employs the k-means clustering algorithm and a supervised cluster-application mapping strategy.

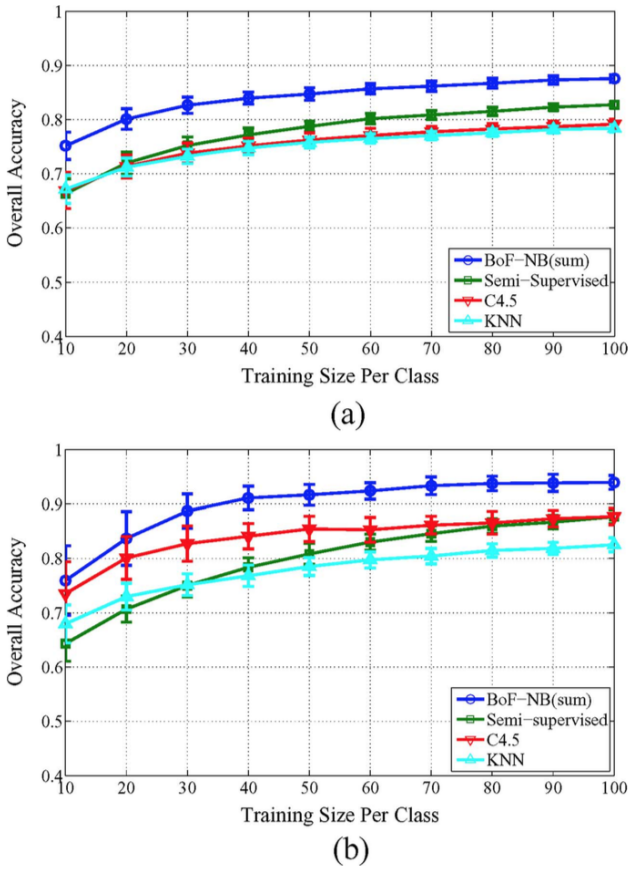
For semi supervised learning method, a significant number of flows will be identified as unknown if small size if supervised training set is available. Fig. 3[1] shows the classification accuracy of the four competing classification methods versus training data size.

Fig. 6 clearly shows that BoF-NB outperforms other three methods.This graph show that BoF-NB can improve the classification accuracy by aggregating correlated NB predictions.BoF-NB approach provides a solution with considerably fewer training time without dropping performance.This is unlike with methods used in [14]-[18].

VII. Intrusion Detection System based on Hidden Markov Models

In this section we discuss about application of Hidden Markov Model for intrusion detection.[19] Hidden Markov Model is a model based on discrete states, where the states are not observable or hidden.Two probability density functions are associated to each hidden state: first provides the probability of transition to another state, the

Fig. 6. Classification accuracy of four methods (a) on isp dataset, (b) on wide dataset.[1]

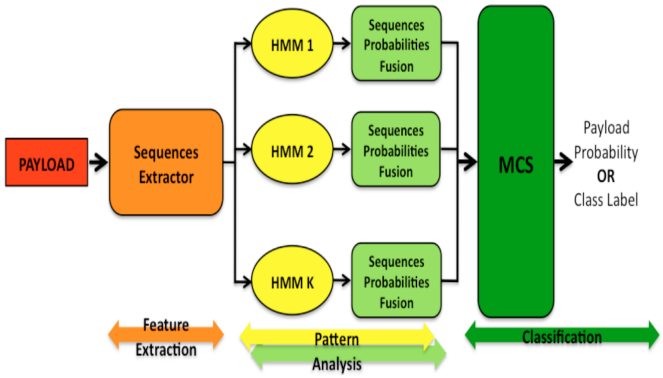


other provides the probability that a given symbol is emitted from that state. This survey covers a method which uses three steps. First step deals with feature extraction. Second step is about pattern analysis. Finally the last step is where actual classification takes place.Fig 7 shows these three steps.

A. *Classifier fusion*

In fig 4 we can clearly see that this method uses multiple classifiers. Statistically it is known that doing this gives us better results.Effectiveness of using multiple classifiers for computer security has been shown in these literatures[20]-[21]. Single result is obtained from these multiple classifiers by calculating the mean of results of all individual classifiers. We will not focus on

Fig. 7. Hidden Markov Model for intrusion detection



Hidden Markov model details in this survey but discuss how it can be applied in intrusion detection in network security.

B. *Functional stages of Hidden Markov Model*

1) Feature Extraction: New attacks appear every day, and a number of variants of known attacks are developed.Due to the dynamic nature of network packets it is very difficult to train a classifier, feature in machine learning becomes paramount. Additionally the features should be chosen by taking into account that it is hard for attackers to craft an attack.

A window of fixed length slides over the payload byte by byte. This group of bytes in single widow is considered as sequence. This window slides across the payload producing multiple sequences in same manner. For example if we consider the payload string *(2,1,2,0,0,1,2,1,0,2)*. Then our process will give out following sequences : *(2-1-2-0-1), (1-2-0-0-1),(2-0-0-1-1),(0-0-1-2-1),(0-1-2-1-0),(1-2-1-0-2).*

2) Pattern Analysis: Output sequences from previous step *(Feature Extraction)* will serve as input to this step. Since Hidden

Markov model is already trained, it knows what normal payload looks like.During the detection test HHM test the probability of any given payload belonging to or being similar to normal payload.

Since we are using multiple classifiers in this approach, all the classifiers are allow to calculate their probabilities first. Then arithmetic mean is calculated of results of all classifiers to get the final result.

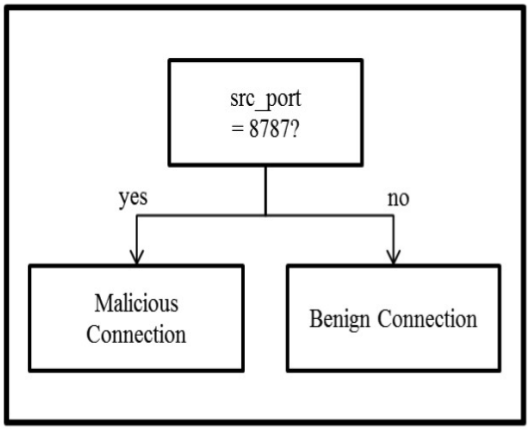
3) Classification: A given payload is classified as normal or attack by comparing its calculated probability to previously assigned threshold probability. Value of threshold probability depends upon the accepted trade off between detection and false positive rates.

VIII. Intrusion Detection Using Decision tree

In the proficiency or experience of the computer network the attacks have increased and it is becoming very difficult to detect and counter intrusions. Combating these new types of attack, we need a variety of tools and technique to detect and defend against attacks. A technique from the data mining field can assist with this kind of task. It provides the identification of the malicious activity and technique to defend against attack.

What is decision Tree? A decision tree is defined as a predictive modelling technique from the field of statistic and Machine learning that builds a simple tree-like structure to the model underlying pattern of data. [22]. Algorithm classification creates a decision tree: Shown in the figure-8 below. To create the tree, it need the information that it gets from identifying patterns from the existing dataset. It creates a simple rule from the pattern they learn from the data, to distinguish between the various data that are pre-classified data set.

Fig. 8. Decision tree



To get a good idea for the classification process, take the previous decision tree as a base model.

Taking an input which is pre-categorized network data is also called features of the data. They are vital quantifiable characteristics of the data. The features can be continuous discrete; these features are stored in form of tables such as database or spreadsheet or in a tabular form. A good example for this is Attribute Relation File Format. Decision tree are built with classification of algorithm, creating an optimal tree computationally infeasible. A lot of greedy algorithm, are based on minimising the entropy associated with different rule set recursively for a quick and effective sub optimal tree. To evaluate our system, we need two major indications. I. TN (Total Normal): total number of normal record. II. TA (Total Attack): total number of attack record. III. Detection Rate = [TA-FN/TA] *100 IV. FP (False Positive): total number of normal record that are indicated as anomalous. V. FN (False Nega-

tive): total number of anomalous record.

Using Decision Tree for Intrusion Detection. To get insight of Intrusion detection can be accomplished using a straight process. Steps for implementation of Decision tree for Intrusion detection.

1. Data Collection and Tool Acquisition: Two prerequisites for the analysis of the data are I. Identify and collect data of interest. II. Identify datamining tool.

2. Feature Extraction: The gathered data required pre-processing phase to convert into form which is necessary for decision tree algorithm.

3. Training Tree: The tree can be trained from the processed data and tool.

4. Analysis of Result: Important step to understand the resulting model and its rule set.

5. Running Decision tree (Rules in Real time): It is using the result of analysis to run the decision rules in real time.

IX. Difficulty in Using Machine Learning for Network Security

Machine learning does provide a lot of difficulty in the general case and it?s even more when it is used for securing the network. As already stated, Machine learning models get better when more training data is available. Even though there are a lot of sources of network data available, there are still a few challenges one encounters when using machine learning for securing a network.

Lack of Ground Truth: It is usually difficult to determine ahead of time what characteristics of a packet that makes it a malicious packet. Unlike other applications of Machine learning, one can easily clarify an event as true or false but this distinction does not immediately jump out when applied to security. Attackers have numerous ways of masking a packet to sure it does not look suspicious. There is also a huge imbalance of data: we may have thousands of successful logins every day and just one 'bad' one. It becomes the task of the machine learning algorithm to engineer to come up with a model that distinguishes that single one.

Disproportionate Cost of False Negative: Even while trying very hard to avoid over-fitting the model, the cost of a single false negative can be millions or billions of dollars or can even deal a huge blow on a company's reputation. It becomes a big challenge for the machine learning model to avoid over-fitting and yet achieve the best possible accuracy.

Constantly changing environment: The network environment is constantly changing with many Servers or devices being installed and uninstalled all the time. Users may also be assigned to a different department and may have their privileges improved or reduced. Different services may also be commissioned or decommissioned and this means the machine learning model must be dynamic to accommodate these changes.

Adversarial setting: An attacker could still deliberately avoid detection by exploring the idea behind machine learning. This can be done by masking malicious packet to appear like normal traffic in other to trick the machine learning model.

X. Conclusion

Conclusively, Machine learning algorithms have shown a lot of promise in its application to security. Its application to network classification, anomaly detection and intrusion detection have shown a high predictive rate. It is important to notice that these algorithms are implemented in such a way that the errors (both false positives and false negatives) are reduced. Support vector machines, Naive Bays predictions and decision tree algorithms have characteristics that can be used to classify network traffic and can be modified to provide more security. There are also some challenges with using ML technique to provide security. The data may be too large since network traffic can change in real time hence the process of classifying the network must be dynamic. The data set may also be small in which case a flow correlation function must be incorporated into the classification.

References

[1] Wang, Guojian, Linmi Tao, Huijun Di, Xiyong Ye, and Yuanchun Shi. "A scalable distributed architecture for intelligent vision system." Industrial Informatics, IEEE Transactions on 8, no. 1 (2012): 91-99.

[2] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. New York: Wiley, 2001.

[3] J.Kittler,M.Hatef,R.Duin,andJ.Matas,"Oncombiningclassifiers," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 3, pp. 226?239, Mar. 1998.

[4] J.Erman,A.Mahanti,M.Arlitt,I.Cohen,andC.Williamson,"Offline/ realtime traffic classification using semi-supervised learning," Performance Evaluation, vol. 64, no. 9-12, pp. 1194?1213, Oct. 2007.

[5] H. Kim, K. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, "Internet traffic classification demystified: Myths, caveats, and the best practices," in Proc. ACM CoNEXT Conf., New York, 2008, pp. 1-12.

[6] Y.-S. Lim, H.-C. Kim, J. Jeong, C.-K. Kim, T. T. Kwon, and Y. Choi, "Internet traffic classification demystified: On the sources of the discriminative power," in Proc. 6th Int. Conf., Ser. Co-NEXT'10, New York, 2010, pp. 9:1-9:12, ACM.

[7] Mahdi Zamani and Mahnush Movahedi, "Machine Learning Techniques for Intrusion Detection". Department of Computer Science University of New Mexico.

[8] Jun Zhang*, Yang Xiang, Member, IEEE, Yu Wang, Wanlei Zhou, "Network Traffic Classification Using Correlation Information."

[9] A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques," in SIGMETRICS Perform. Eval. Rev., Jun. 2005, vol. 33, pp. 50-60.

[10] S. Zander, T. Nguyen, and G. Armitage, "Automated traffic classification and application identification using machine learning," in Proc. Ann. IEEE Conf. Local Computer Networks, Los Alamitos, CA, 2005, pp. 250?257.

[11] J. Erman, M. Arlitt, and A. Mahanti, "Traffic classification using clustering algorithms," in Proc. SIGCOMM Workshop on Mining Network Data, New York, 2006, pp. 281-286.

[12] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian neural networks for internet traffic classification," IEEE Trans. Neural Netw., vol. 18, no. 1, pp. 223-239, Jan. 2007.

[13] M.Roughan,S.Sen,O.Spatscheck,andN.Duffield,"Class-of-service mapping for QoS: A statistical signature-based approach to IP traffic classification," in Proc. 4th ACM SIGCOMM Conf. Internet Measurement, New York, 2004, pp. 135-148.

[14] S. Zander, T. Nguyen, and G. Armitage, "Automated traffic classification and application identification using machine learning," in Proc. Ann. IEEE Conf. Local Computer Networks, Los Alamitos, CA, 2005, pp. 250-257.

[15] J. Erman, M. Arlitt, and A. Mahanti, "Traffic classification using clustering algorithms," in Proc. SIGCOMM Workshop on Mining Network Data, New York, 2006, pp. 281-286.

[16] L. Bernaille, R. Teixeira, I. Akodkenou, A. Soule, and K. Salamatian, "Traffic classification on the fly," in Proc. SIGCOMM Comput. Commun. Rev., Apr. 2006, vol. 36, pp. 23-26.

[17] Y. Wang, Y. Xiang, and S.-Z. Yu, "An automatic application signature construction system for unknown traffic," Concurrency Computat.: Pract. Exper., vol. 22, pp. 1927-1944, 2010.

[18] A. Finamore, M. Mellia, and M. Meo, "Mining unclassified traffic using automatic clustering techniques," in Proc. TMA Int. Workshop on Traffic Monitoring and Analysis, Vienna, Austria, Apr. 2011, pp. 150-163.

[19] Davide Ariu, Roberto Tronci, and Giorgio Giacinto, "An Intrusion Detection System based on Hidden Markov Models," Department of Electric and Electronic Engineering, University of Cagliari Piazza d?Armi, 09123 Cagliari, Italy

[20] L. Kuncheva, Combining Pattern Classifiers, Wiley, 2004.

[21] I. Corona, G. Giacinto, C. Mazzariello, F. Roli, C. Sansone, Information fusion for computer security: State of the art and open issues, Information Fusion 10 (2009) 274?284.

[22] Bloedorn, E., Christiansen, A. D., Hill, W., Skorupka, C., Talbot, L. M., Tivel, J. (n.d.). Data Mining for Network Intrusion Detectio

[23] Tich Phuoc Tran, Pohsiang Tsai, Tony Jan,Xiaoying Kong. "Network Intrusion Detection using Machine" Centre for Innovation in IT Services and Applications (iNEXT).University of Technology, Sydney.

[24] Jun Zhang, Xiao Chen, Yang Xiang, Wanlei Zhou, Jie Wu, "Robust Network Traffic Classification"