International Organization of Research & Development (IORD)
ISSN: 2348-0831
Vol 10 Issue 01 | 2022

Network Based Classification Using Machine Learning and Deep LearningAlgorithm

Ms. Shivani Ganesh Bonde
Assistant Professor, Computer Science and Engineering Department, SGBAU, Amravati University

DOI: 10.5281/zenodo.7161992

ABSTRACT

More complexity of internet increases with increase in its use. And to handle this increase network facility many challenges have been generated within the system and to cope with this situation many networks models has been designed and upgraded time to time. Some approaches such as Software Defined Networking (SDN) is helpful to provide centralized mechanism building strategy to handle this ever increasing demand. Machine learning concept has been evolved to manage the network administration situation and helps to measure the network essential requirements to minimize the burden over the system. To understand this scenario the network prediction or network traffic prediction system was developed bymaking use of some ML process and make it easy to manage the resources in an efficient way.

Keywords— Network measurement, software defined networking, machine learning, deep learning,

1. INTRODUCTION

The fast development of the web and communication devices has created bigger and more complicated network structures, adapting and developing bigger hubs, routers, switches etc. This complexity in networks has introduced an overflow of vast amounts of traffic data and contributed to the challenges in network management and traffic optimization, including traffic measurement (e.g. traffic classification) and traffic prediction. In parallel, we are seeing two promising solutions to assist manage networks more efficiently: SDN and Machine Learning. SDN provides a centralized access and control mechanism to all or any networking devices, where the SDN controller cant only monitor and measure all kinds of network parameters and metrics, but may also make a more informed and efficient decision about resources allocation and routing, since its a world view of everything within the network. However, the knowledge of data an SDN may be overwhelming. While the SDN controller itself can be made scalable, for instance by running it during a cloud, still efficient algorithms are needed to extract the desired measurements and knowledge from the received data. Here is where Machine Learning can help. Many of the traffic classification and traffic prediction issues can be performed efficiency by various ML algorithms, improving the system performance while maintaining relative simplicity in design. During this survey, we review existing approaches for traffic classification and traffic prediction which use ML in an SDN context. We especially target on ML's subcategory of Deep Learning (DL), which is usually not covered within the existing surveys. Therefore, our contribution is covering DL methods for traffic prediction, which is mostly not covered in the existing surveys, while we also cover some newer works in ML and DL for both traffic classification and traffic prediction that existing surveys have not covered. Finally, we investigate open research issues and suggest possible future research avenues. [1]
The following are the ML and DL algorithms utilize in this survey. Note that all but the last one are supervised.

- Nearest Centroid (NC)
- Naive Bayes (NB)
- Decision Tree (DT)
- Random Forest Tree (RF)
- Support Vector Machine (SVM)
- Multi-Class Support Vector Machine(MCSVM)
- Laplacian Support Vector Machine(LapSVM)
- Adaptive Boosting (AdaBoost)
- Gradient Adaptive Boosting (G-AdaBoost)
- M5Rules
- Linear Regression
- Polynomial Regression:
- K-means:

1.1 Traditional Deep Learning Algorithms

Deep Learning uses multiple layered neural networks which are biologically-inspired computing systems with input, hidden and output layers consisting of interconnected neuron-like nodes. The nodes contain activation functions. Information is fed through the input layer. The pattern recognition process is done in the hidden layer via activation functions and the answer is presented in the output layer. Each layer takes the output of the previous layer(s) as input and applies non-linear transformation to extract useful features for classification.

- Convolutional Neural Network (CNN)
- Autoencoders (AE)
- Recurrent Neural Network (RNN)

2. RELATED WORK

In remote sensing image classification, we are usually given a reduced set of labeled samples to develop the classifier. Supervised classifiers like as support vector machines (SVMs) excel in using the labeled information, being(regularized) maximum margin classifiers also equipped with an appropriate loss function. These methods, nevertheless, have to use the knowledge to contained within the wealth of unlabeled samples, which is thought as semisupervised classification. In semi- supervised learning (SSL), the algorithm isgiven with some available supervised information additionaly to the unlabeled data. The framework of semi-supervised learning is incredibly active and has recently attracted a substantial amount of research. [2] The traditional method of network traffic classification relies on the network port implementation. Through the network protocol and application software for data transmissionwhen the final port requirements, the portnumber and also the specific network protocol, application software one by one to see the utilization of the port of the network traffic to complete the identification of network traffic. However, there are many new applications (Streaming, Gaming and P2P) appearing inadditionaly to traditional applications, the port- based protocol identification method isn't any longer reliable. The classification method supported onfeature field recognition is to pre- establish the appliance layer identification rule base for network traffic generated by each network application. It is to ascertain the applying layer identification rule base of the network traffic generated by each network application ahead, to rule out the info content of the information to be identified, and to see the applying form of network traffic per the matching result. However, the classification method based on feature field recognition canonly identify the existing traffic type, but also can not contribute to the encrypted data.[3] Traffic classification is an important task in modern communication networks. Due to the rapid growth of high-throughput traffic demands, to properly manage network resources, it is vital to recognize different types of applications utilizing network resources. Consequently, accurate traffic classification has become one of the prerequisites for advanced network management tasks such as providing appropriate Quality-of-Service (QoS), anomaly detection, pricing, etc. Traffic classification has attracted a lot of interests in both academia and industrial activities related to network management. [4] Network Traffic classification has strained important consideration over the past few years. Classifying traffic flows by their generation applications plays very essential task in network security and management, such as, lawful interception and intrusion detection, Quality of Service (QoS) control. Conventional traffic classification methods [1] include the port-based prediction methods and payload-based deep inspection methods. In current network environment, the standard methods suffer from a variety of practical problems such as dynamic ports and encrypted applications. Recent research efforts are absorbed on the appliance of machine learning techniques to traffic classification constructed on flow statistical features. It can instinctively explore for and describe practical structural patterns in a very supplied traffic dataset, which is useful to logically conduct traffic classification. The flow statistical feature founded traffic classification can be understood by using supervised classification algorithms or unsupervised classification (clustering) algorithms.[5] Traffic classification methods are flow and packet based measurements have been previously researched using various techniques starting from automated machine learning (ML) algorithms to deep packet inspection (DPI) for accurate application identification. Port and protocol analysis, once the default method for traffic identification is now considered obsolete as most applications use dynamic ports, employ HTTPS or encrypted SRTP or use tunneling which makes classification near impossible. Deep packet inspection (DPI) is useful; however the computational overhead and extra hardware required for packet analysis severely limit its practical implementation for network operators. Moreover, aggregation based traffic monitoring techniques using flow measurements have proliferated in recent years because of their inherent scalability and easy of implementation similarly as well as compatibility with existing hardware using standardized export formats such as NetFlow and IPFIX [6].

3. ANALYSIS OF PROBLEM

Neural network centric behavior during training process overheads the repetition of information itself. Availability of labeled data decreases accuracy in classification of assorted algorithms. Finding how to mix unsupervised learning with supervised learning to show NN the way to learn with fewer data may be promising area of research. Furthermore, teaching NN to accumulate its knowledge will make it more practical and efficient in learning new things, andthus, less data are required for training. and extensibility. However, the decoupling the knowledge of and control planes has also made the network more at risk of security issues. For instance, since the network is managed by a one controller, overloading it with malicious flows creates a challenging problem. To deal with this problem, DL algorithms may be used more often in detecting suspicious flows and anomaly based attacks.

3.1 Traffic Flow

Network traffic involves encrypted/ encapsulated flow packets which hide the features of the flows. Classifying such traffic requires advanced DL methods that may reveal hidden patterns. The evolution within the networking architecture has brought flexibility and extensibility. However, the decoupling the knowledge of and control planes has also made the network more at risk of security issues. For instance, since the network is managed by a one controller, overloading it with malicious flows creates a challenging problem. To deal with this problem, DL algorithms may be used more often in detecting suspicious flows and anomaly based attacks.

3.2 Traffic Certainty

There are various studies on application traffic classification. The foremost common classification method among the assorted methods could be a payload signature based classification method. This methodextracts a unique signature for each payload of each application traffic and classifies it based on the extracted singular signature. Such classification based on payload signature is difficult to apply if the data part is encrypted [8] .

4. PROPOSED WORK

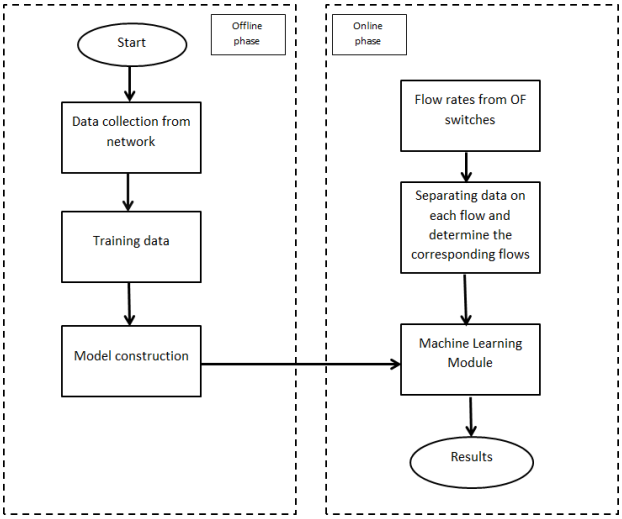

Fig.1 Proposed System Architecture

In the opening, collect application traffic. Duringthis step, we capture from various applications through Microsoft Network Monitor 3.4. Application traffic should be captured in appropriate amount. If the number of traffic isn't adequate, it will not only give accurate classification results, but also will also take an extended time. Real-time traffic classification has the potential to resolve difficult network management problems for Internet service providers (ISPs) and their equipment vendors. Network operators must know what is flowing over their networks promptly in order to that they can react quickly in support of their various business goals. Traffic classification is also a core part of automated intrusion detection systems, used to detect patterns indicative of denial of service attacks, trigger automated re-allocation of network resources for priority customers, or identify customer use of network resources that in way contravenes the operator's terms of service. More recently, governments are also clarifying ISP obligations with relevance to 'lawful interception' (LI) of IP data traffic. Even as telephone companies must support interception of telephone usage, ISPs are increasingly subject to government requests for information on network use by particular

individuals atparticular points in time. IP traffic classification is an integral part of ISP-based LI solutions.[6] diverse. In order that application traffic patterns are getting even more diverse and complicated Inthis situation, network traffic monitoring and analysis is essential for effective network operation and stable service provision. In order to analyze the network traffic, application traffic classification method must be preceded. The proposed system helps to see the actual fact of network congestion and for the answer solution in coordination with system [7]

5. APPLICATION

A. Network Traffic Analysis Using Packet Captures
B. Monitoring The Flow of Traffic With Netflow
C. Detect Application and Protocols in Use
D. Track Bandwidth Usage to Find Bandwidth Hogs
E. Create Custom Reports

6. CONCLUSION

With the rapid Internet volume growth, network environment is daily becoming complex and diverse. In order that application traffic patterns are getting even more diverse and complicated In this situation, network traffic monitoring and analysis is essential for effective network operation and stable service provision. In order to analyze the network traffic, application traffic classification method must be preceded.

The proposed system helps to see the actual fact of network congestion and for the answer solution in coordination with system.

7. REFERENCES

[1]. Machine Learning and Deep Learning Based Traffic Classification and Prediction in Software Defined Networking Ays¸e Rumeysa Mohammed, Shady A. Mohammed, and Shervin Shirmohammadi IEEE Instrumentation and Measurement Society, 978-1-7281-1273-2/19 ©2019 IEEE

[2]. L. G o´mez-Chova, G. Camps-Valls, J. Munoz- Mari, and J. Calpe, "Semisupervised image classification with laplacian support vector ma- chines," IEEE Geoscience and Remote Sensing Letters, vol. 5, no. 3, pp. 336–340, 2008.

[3]. First International Conference on Advanced Algorithms and Control Engineering IOP Publishing IOP Conf. Series: Journal of Physics: Conf. Series 1087 (2018) 062021 doi :10.1088/1742- 6596/1087/6/062021 Network Traffic Classification Based on Deep Learning Jun Hua SHU1, Jiang JIANG2, Jing Xuan SUN3

[4]. Soft Computing https://doi.org/10.1007/s00500-019- 04030-2 METHODOLOGIES AND APPLICATION Deep packet: a novel approach for encrypted traffic classification using deep learning Mohammad Lotfollahi1 · Mahdi Jafari Siavoshani1 · Ramin Shirali Hossein Zade1 ·Mohammdsadegh Saberian1

[5]. Reference 4

[6]. Jamuna .A, Vinodh Ewards S.E / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 2, March -April 2013, pp.1324-1328 1324 Efficient Flow based Network Traffic Classification using Machine Learning Jamuna .A*, Vinodh Ewards S.E**