State of the Art Review of Network Traffic Classification based on Machine Learning Approach

Pallavi Singhal
L.M.College Of Science and Technology
A-Sector,ShastriNagar Jodhpur

Rajeev Mathur, Ph.D.
L.M.College Of Science and technology
A-Sector,ShastriNagar Jodhpur

Himani Vyas
L.M.College Of Science and technology
A-Sector,ShastriNagar Jodhpur

ABSTRACT

Network traffic classification is extensively required mainly for many network management tasks such as flow prioritization, traffic shaping/policing, and diagnostic monitoring. Similar to network management tasks, many network engineering problems such as workload characterization and modeling, capacity planning, and route provisioning also benefit from accurate identification of network traffic .This paper presents review on all the work done related to Network Traffic Management since 1993 to 2013 in various fields like artificial intelligence, neural network, ATM and wireless networks.

General Terms

Network, Network Traffic, Supervised, Unsupervised Learning, Semi-Supervised, Clustering, Classification.

Keywords

Network Traffic, Supervised, Unsupervised Learning, Semi-Supervised, Clustering, Classification, Machine Learning

INTRODUCTION

The earliest work on network traffic classification was done by Gebele et al[1] where they attempted to provide a set of benchmarks for GUI (graphical user interface) applications developed under the X Window System, by proposing a method of classifying the GUI characteristics into classes based on a measure of graphical complexity.Then Gupta et al[2] worked on traffic classification for round – robin scheduling schemes in ATM net To improve the utilization of network resources and facilitate management and control, source types are organized into traffic classes. works.

Kubbar et al[3] also worked on traffic classification and resource allocation in ATM networks. They proposed a technique which simplifies the existing traffic spectrum, and requires simple protocols which guarantee at least an acceptable level of QoS for each application.Barry et al[4] presented a novel multifractal approach to the classification of self-affine network traffic . The fundamental advantages of using multifractal measures include their boundedness and a very high compression ratio of a signature of the traffic, thereby leading to faster implementations, and the ability to add new traffic classes without redesigning the traffic classifier.

Later work on Network Traffic Classification using Dynamic state classifier was carried out by Yephantong et al[5]. where the machine learns to construct new states while simultaneously performing its normal operations. New state construction occurs every time a new event signature is received, and the set of weighted final states are adjusted accordingly.

Correct classification of traffic flows according to the application layer protocols that generated them is essential for most network-management, resource allocation and intrusion detection systems in TCP/IP networks. mechanisms based on full payload analysis are too computationally demanding to be run on most high-bandwidth links. Hence Crotti et al[6] gave a statistical approach to IP-level classificaton of network traffic.Then Kwangjin et al[7] gave the method for classification of network trafficwhich used pattern matching in packet payload as well as application port number.Gyu myoung Lee[8] discussed techniques like multi-path routing using traffic splitting, constraint-based routing, path protection etc. Multipath routing is capable of aggregating the resources of multiple paths and reducing the blocking capabilities in quality of service (QoS) oriented networks, allowing data transfer at higher rate when compared to single path. It also increases the reliability of delivery.

Kiziloren,T and German,E[9] introduced a classification method for analyzing network traffic behavior. The measurement of traffic is performed by using Simple Network Management Protocol (SNMP).They proposed a , Self Organizing Maps (SOM)-based classifier to discriminate three types of network traffic as port scanning, heavy-download and the rests.Auld,T[10] presented a traffic classifier that can achieve a high accuracy across a range of application types without any source or destination host-address or port information.They used supervised machine learning based on a Bayesian trained neural network.

Balagani et al[11] presented 'D-CAD,' a novel divergence-measure based classification method for anomaly detection in network traffic. The D-CAD method identifies anomalies by performing classification on features drawn from software sensors that monitor network traffic.Then Liu Yingqiu et al[12] introduced the different levels in network traffic-analysis and the relevant knowledge in machine learning domain, analysis the problems of port-based and payload-based methods in traffic classification. They experimented with unsupervised K-means to evaluate the efficiency and performance.According to their reseach the experimental results on different datasets convey that the method can obtain up to 80% overall accuracy, and, after a log transformation, the accuracy is improved to 90% or more.

Then later Rajkamal et al[13] worked on Packet classification for Network Processors in WSN(Wireless sensor networks).In their paper they discussed the design of co-processor which could assist the packet classification for NPs. For this work they used neural network to classify the packets based on nature of the data to overcome the traffic in WSN.Raahemi et al[14] classified Peer-to-Peer traffic using Fuzzy ARTMAP. Fuzzy ARTMAP is an incremental learning classifier suitable for mining stream of data. . They captured Internet traffic at a main gateway router, performed pre-processing on the data, selected the most significant attributes, and prepared a training data set to which the fuzzy ARTMAP algorithms were applied.

Network traffic policy verification is the analysis of network traffic to determine if the observed traffic is in compliance or violation of the applied policy.Thus finding the best possible learning model in combination with extracting the best possible feature-set is a necessary requirement to design accurate traffic classification models.Hence Teufl,P et al[15] in their work presented a framework to simplify the empirical model selection and feature extraction process.

Then a step forward was taken by Becker,M et al[16] in the field of sensor networks. In their work they evaluated the feasibility of both classical machine learning algorithms and bio-inspired algorithms for misbehavior detection in sensor networks.This kind of data analysis was not done in previous studies. And they concluded that classical algorithms have equal or even better detection capabilities compared to some bio-inspired algorithms.

One of the most important applications of traffic classification is in the field of network security.Dainotti,A[17] proposed a packet-level traffic classification approach based on Hidden Markov Model (HMM). Classification is performed by using real network traffic and estimating - in a combined fashion - Packet Size (PS) and Inter Packet Time (IPT) characteristics, thus remaining applicable to encrypted traffic too. Accurate and efficient network traffic classification is an important network management task. Two way messages in a session follow the underlying application protocol to exchange information. In their paper, Gang Shen and Lian Fen[18] proposed a novel application classification method based on message statistics, concisely representing the protocols' unique characteristics.Then in the same field Ai-min Yang et al[19] proposed a P2P network traffic classification method using SVM classifier. . By their method, the P2P network traffic can been classified according to application types with statistical characteristics of network traffic. . Their paper introduced the network traffic classification problem on four application types of P2P, namely, BitTorrent, PPLive, Skype, MSN.

A research was carried out by Yizhen Liu et al[20]. Their paper discussed an efficient hybrid packet classification in gigabits traffic control systems using second-generation programmable network processor. They addressed the problem of inaccurate packet classification and analyze the payload of applications, presented the packet classification using not only packet header but the first 64-bit payload. And finally described the software pipeline architecture and hardware design for our approach with network processor. A new approach based on the implementation of artificial neural network ensemble with the error-correcting output codes (ECOC) was proposed by Xiao Xie et al[21] for classification of multi-class network traffic.

Then Satoh,A et al[22] proposed Traffic Classification in Mobile IP Networks. In their paper, they proposed sub-flow selection with application behaviors, and the method solves a critical problem of how to select appropriate sub-flows for achieving traffic classification in mobile IP network. Although at a very early stage of development, they proposed method which shows promising preliminary results through the experiments on a reduced set of applications.

To meet the requirements of the network activities and take into account P2P traffic classification challenges, a promising method is to use Machine Learning (ML) techniques and identify network applications based on flow features. For this Chengjie Gu and Shunyi Zhuang[23] presented a novel P2P traffic identification approach using back propagation neural network.

In their paper ,Gargiulo,F and Sansone,c[24] proposed to apply an algorithm for finding out and cleaning mislabeled training sample in an adversarial learning context, in which a malicious user tries to camouflage training patterns in order to limit the classification system performance.

In order to tackle the problem of P2P data encryption persecuting P2P identification, in their paper,Gi Yiran and Wang Suoping[25] proposed a traffic identification method for specific P2P based on multilayer tree combination classification BP-LVQ Neural-Network baseing itself on traffic characters of P2P. This method improved the P2P identification with BP Neural-Network, by abstracting attributes of P2P flow statistics, selecting the optimal attribute subset, establishing a P2P classifier through the multilayer combination with BP Neural-Network and LVQ Neural-Network.

Computer networks became a ubiquitous part of modern society. As the spread of networks continued to increase, so do the various applications for the underlying technology. Thus traffic classification became and remains important to network administrators. In their paper,Sanders,S et al[26] used preliminary results for multi-dimensional piecewise polynomials to model network traffic. Machine learning based techniques gained more and more attentions in the next few years.Runyan Sun et al[27] used a distributed host based traffic collection platform (DHTCP) to gather traffic samples with accurate application information on user hosts.

Then Wang Ruoyu et al[28] proposed a new re-sampling method named tuning sampling for supervised machine learning (SML) to ease the problem of data skew in internet traffic classification. And compared it with uniform sampling and stratified sampling methods using C4.5 classification algorithm.

Later Abdennebi,M et al[29] in their paper proposed an adaptive Call Admission Control for IEEE 802.16 scheduled flows. They took advantage of the variability of traffics and adapted the CAC according to the characteristics of incoming flows.

Toplak,W et al[30] proposed Novel Road Classifications for large scale traffic networks. Their paper aimed to improve the scalability of link travel time predictions by combining information from roads with similar characteristics.Thus establishing a highly sophisticated large-scale Traffic Information System (TIS) for large road network.

A study on Process of Network Traffic Classification Using Machine Learning was carried out later by Jian-Min Wang et al[31]. . In their paper, a detailed workflow of machine learning based network traffic classification in campus network of SunYat-sen University is described, including steps such as data preparation and model construction. They also performed simple experiments to prove the effectiveness of machine learning approach.

Shrivastav,A.andTiwari,A[32].workedonNetworkTrafficClassificationUsingSemiSupervisedApproach. This traffic classification methodology uses only flow statistics to classify traffic. The approach consists of two steps, clustering and classification. Clustering partitions the training data set into disjoint groups. After making clusters, classification is performed in which labeled data are used for assigning class labels to the clusters. A

KDD Cup 1999 data set is being taken for testing this approach. It includes many kind of attack data, also includes the normal data. The testing results are then compared with SVM based classifier.

HuTinget al[33] proposed a network traffic classification method based on Kernel Self Organizing Maps(KSOM) , which replaces Euclidean distance with non-Euclidean distance induced by kernel function, and adopts it to estimate the matching degree between the input pattern and the connection weight.

Ciresan, D et al[34] proposed a committee of neural networks for traffic sign classification. An efficient SVM(Support Vector Machine)-based method for multi-classnetwork traffic classification was given by Ning jing et al[35]. They proposed a novel scheme for SVM-based traffic classification (called fuzzy tournament).

Then later work was carried out on volunteer-based system for classification of traffic in computer networks by Bujlow et al[36]. They developed a new system in which the data are collected from client machines. Their paper presented design of the system, implementation, initial runs and obtained results. Furthermore, it proves that the system is feasible in terms of uptime and resource usage, assesses its performance and proposes future enhancements.

Zaman,B et al[37] proposed Implementation vehicle classification on Distributed Traffic Light Control System neural network based. Distributed Traffic System Control System is a real-time adaptive traffic light system with traffic condition for minimize the probability of traffic congestion.
A Parallelized Network Traffic Classification Based on Hidden Markov Model was proposed by Xuefeng Mu and Wenjun Wu[38]. Their paper implemented a network traffic classification method on the basis of Guassian Mixture Model-Hidden Markov Model using packet-level properties in network traffic flows (PLGMM-HMM).

QoS routing algorithm based on traffic classification in LEO satellite networks was proposed by Jiang Wenjuan and Zong Peng[39] . Low Earth Orbit (LEO) satellite networks are expected to provide a variety of multimedia applications. In order to satisfy different QoS requirements and optimize utilization of network resources, a novel traffic classification routing algorithm (TCR) was proposed. The key technique of TCR is investigating the traffic classification link-cost metrics (TCM) for different traffic classes, which selects the next-hop based on real time link state information. The advantages of the proposed scheme, in terms of average path delay and blocking probability, are corroborated by ample simulations, where significant gains in performance are achieved.

Wengang Zhou et al[40] gave a new approach based on feed-forward neural network for accurate traffic classification, which eliminates the disadvantages of port-based or payload-based classification methods.

Internet growth drives the emergence and usage of new applications, which have specific requirements for Quality of Service (QoS). Thus, differentiation of packets which belong to certain QoS classes becomes a crucial factor to ensure the desirable quality for the applications. Hence Gomes et al[41] proposed an agent for real-time traffic classification based on QoS classes for virtual network environments, which aims at forwarding the flow to the adequate virtual network according to the QoS class. The study was made using the Mininet emulator and the Openflow Protocol.
Haihong Gao et al[42] proposed Traffic classification and observer design of cable networks . In their paper they proposed an observer model and design for cable networks. The observer design is recast into an online traffic classification problem. Using dynamic network traffic phases observed through simulation, the network classifier is designed with a hidden Markov model to determine the network states.

A method for classification of network traffic based on C5.0 Machine Learning Algorithm was proposed by Bujlow, T et al[43] . To overcome the drawbacks of existing methods for traffic classification, usage of C5.0 Machine Learning Algorithm (MLA) was proposed. They collected accurate traffic data, presents arguments used in classification process, introduced the C5.0 classifier and its options, and finally evaluated and compared the obtained results.

Then Jun Zhang et al[44] presented a new semi-supervised method to effectively improve traffic classification performance when few supervised training data are available. They proposed to incorporate flow correlation into both training and testing stages.

Recently, the need of traffic classification and applications identification has attracted numerous research efforts. Based on statistical attribute analysis, recent research studies employed machine learning algorithms for building traffic classifiers. The machine learning based traffic classification achieves high accuracy and becomes prominent scheme. Nen-Fu Huang et al[45] in their paper proposed the framework of cloud-based traffic classification service for sharing model and parallel classification.

Zhang, Jun et al[46] recently proposed Network Traffic Classification Using Correlation Information.. They proposed a novel non parametric approach for traffic classification, which can improve the classification performance effectively by incorporating correlated information into the classification process.They analyzed the new classification approach and its performance benefit from both theoretical and empirical perspectives.

REFERENCES

[1] Gebele, A.J.; Khalil, K.,1993. Network traffic workload classification methods for workstation GUI applications

[2] Gupta, S.; El Zarki, M,1993, Traffic classification for round-robin scheduling schemes in ATM networks

[3] Kubbar, O.; Mouftah, H.T.,1996, Traffic classification and resource allocation in ATM networks

[4] Barry, R.L.; Kinsner, W.2004, Multifractal characterization for classification of network traffic.

[5] Yeophantong, T.; Pakdeepinit, P.; Moemeng, P.; Daengdej, J.,2005, Network Traffic Classification Using Dynamic State Classifier

[6] Crotti, M.; Gringoli, F.; Pelosato, P.; Salgarelli, L.,2006, A statistical approach to IP-level classification of network traffic

[7] Kwangjin Choi; Jun Kyun Choi,2006, Pattern Matching of Packet Payload for Network Traffic Classification.

[8] Gyu Myoung Lee; Jun Kyun Choi,2006, Multipath Routing for Traffic Engineering with Flow Classification in GMPLS Network.

[9] Kiziloren, T.; Germen, E.,2007, Network traffic classification with Self Organizing Maps.

[10] Auld, T.; Moore, A.W.; Gull, S.F.,2007, Bayesian Neural Networks for Internet Traffic Classification.

[11] Balagani, K.S.; Phoha, V.V.; Kuchimanchi, G.K.,2007, A Divergence-measure Based Classification Method for Detecting Anomalies in Network Traffic.

[12] Liu Yingqiu; Li Wei; Li Yunchun,2007, Network Traffic Classification Using K-means Clustering.

[13] Rajkamal, R.; Vanaja Ranjan, P.,2008, Packet classification for Network Processors in WSN traffic using ANN.

[14] Raahemi, B.; Kouznetsov, A.; Hayajneh, A.; Rabinovitch, P.,2008, Classification of Peer-to-Peer traffic using incremental neural networks (Fuzzy ARTMAP).

[15] Teufl, P.; Payer, U.; Amling, M.; Godec, M.; Ruff, S.; Scheikl, G.; Walzl, G.,2008, InFeCT - Network Traffic Classification

[16] Becker, M.; Bohlmann, S.; Schaust, S.,2008, Traffic analysis and classification with bio-inspired and classical algorithms in sensor networks.

[17] Dainotti, A.; de Donato, W.; Pescape, A.; Salvo Rossi, P.,2008, Classification of Network Traffic via Packet-Level Hidden Markov Models.

[18] Gang Shen; Lian Fan,2008, Network Traffic Classification Based on Message Statistics

[19] Ai-min Yang; Sheng-yi Jiang; He Deng,2008,A P2P Network Traffic Classification Method Using SVM.

[20] Yizhen Liu; Daxiong Xu; Zhixin Mu; Jiayi Qin,2009, Efficient Hybrid Packet Classification in Traffic Control System Using Network Processors.

[21] Xie; Bo Yang; Yuehui Chen; Lin Wang; Zhenxiang Chen,2009, Network Traffic Classification Based on Error-Correcting Output Codes and NN Ensemble.

[22] Satoh, A.; Osada, T.; Abe, T.; Kitagata, G.; Shiratori, N.; Kinoshita, T.,2009, Traffic Classification in Mobile IP Network.

[23] Chengjie Gu; Shunyi Zhuang,2010, A novel P2P traffic classification approach using back propagation neural network.

[24] Gargiulo, F.; Sansone, C.,2010, Improving Performance of Network Traffic Classification Systems by Cleaning Training Data.

[25] Gu Yiran; Wang Suoping,2010, Traffic Identification Method for Specific P2P Based on Multilayer Tree Combination Classification by BP-LVQ Neural-Network.

[26] Sanders, S.; Fairbanks, K.; Jampana, S.; Owen, H.,2010, Visual network traffic classification using multi-dimensional piecewise polynomial models.

[27] Runyuan Sun; Bo Yang; Lizhi Peng; Zhenxiang Chen; Lei Zhang; Shan Jing,2010, Traffic classification using probabilistic neural networks.

[28] Wang Ruoyu; Liu Zhen; Zhang Ling,2010, A new re-sampling method for network traffic classification using SML.

[29] Abdennebi, M.; Ghamri-Doudane, Y.; Yan Li,2010, Adaptive CAC with traffic flows classification for IEEE 802.16 networks.

[30] Toplak, W.; Koller, H.; Dragaschnig, M.; Bauer, D.; Asamer, J.,2010, Novel road classifications for large scale traffic networks.

[31] Jian-Min Wang; Cheng-Lu Qian; Chun-Hui Che; Hai-Tao He,2010, Study on Process of Network Traffic Classification Using Machine Learning.

[32] Shrivastav, A.; Tiwari, A.,2010, Network Traffic Classification Using Semi-Supervised Approach.

[33] Hu Ting; Wang Yong; TaoXiaoling,2010,Network traffic classification based on Kernel Self-Organizing Maps.

[34] Ciresan, D.; Meier, U.; Masci, J.; Schmidhuber, J.,2011, A committee of neural networks for traffic sign classification.

[35] Ning Jing; Ming Yang; Shaoyin Cheng; Qunfeng Dong; Hui Xiong,2011, An efficient SVM-based method for multi-class network traffic classification.

[36] Bujlow, T.; Balachandran, K.; Riaz, T.; Pedersen, J.M.,2011, Volunteer-based system for classification of traffic in computer networks.

[37] Zaman, B.; Jatmiko, W.; Wibowo, A.; Imah, E.M.,2011, Implementation vehicle classification on Distributed Traffic Light Control System neural network based.

[38] Xuefeng Mu; Wenjun Wu,2011, A Parallelized Network Traffic Classification Based on Hidden Markov Model.

[39] Jiang Wenjuan; Zong Peng,2011, QoS routing algorithm based on traffic classification in LEO satellite networks.

[40] Wengang Zhou; Leiting Dong; Bic, L.; Mingtian Zhou; Leiting Chen,2011, Internet traffic classification using feed-forward neural network.

[41] Gomes, R.L.; Mauro Madeira, E.R.,2012, A Traffic Classification Agent for Virtual Networks Based on QoS Classes.

[42] Xue Han; Yiqing Zhou; Liang Huang; Lin Han; Jinlong Hu; Jinglin Shi,2012, Maximum entropy based IP-traffic classification in mobile communication networks.

[43] Bujlow, T.; Riaz, T.; Pedersen, J.M.,2012, A method for classification of network traffic based on C5.0 Machine Learning Algorithm.

[44] Jun Zhang; Chao Chen; Yang Xiang; Wanlei Zhou,2012, Semi-supervised and Compound Classification of Network Traffic.

[45] Nen-Fu Huang; Gin-Yuan Jai; Chih-Hao Chen; Han-Chieh Chao,2012, On the cloud-based network traffic classification and applications identification service.

[46] Zhang, Jun; Xiang, Yang; Wang, Yu; Zhou, Wanlei; Xiang, Yong; Guan, Yong,2013, Network Traffic Classification Using Correlation Information.