Role of Machine Learning Applications in Communication Network's Traffic Classification to Manage TCP Congestion

[1]Nasir Jamal, [2]M Zulqarnain, [3]M Waqas Boota, [4]S M Mohsin, [5]Javed A. S

[1, 2, 3, 4, 5] Department of Computer Science, Virtual University of Pakistan, Lahore

Abstract: The examination of packet flow in network application using port number and payload-based identification is important area of network surveillance and management. Machine learning is alternative approach to identify network applications using payload independent features inter-arrival time and packet length. The decision trees like Bayesian Network, Naïve bayes, C4.5 Naïve Bayes Tree algorithm used to demonstrate the impact on performance of reduction of feature set. The computational performance based on algorithms for traffic classification differs from classifying accuracy alone. In wireless and wired heterogeneous networks TCP is inadequate and react to losses due to link failure and congestion in same way. To distinguish packet loss causes, proposing a TCP imparting with a classifier to improve its performance in wireless and wired networks since not able to control error's and congestion in TCP while congestion control ignore the packet losses due to link error and initiate method to recover packets only losses due to congestion. Pre-classified packet losses gathered in a database used for classification algorithm using supervised learning to classify losses. Through of TCP with this classifier improves in both wired and wireless networks. This algorithm was introduced for the integration of OLSR and delay prediction. Since we use Neural network with OSLR (OSLR_NN) that can reduces delay in heavy traffic and improve the data packet delivery ratio. Our work shows the role of neural network in delay prediction.

Keywords: Transmission Control Protocol (TCP), User Datagram Protocol (UDP) Neural Network (NN), Quality of Service (QoS).

1. INTRODUCTION

In internet transmission control protocol is used at large scale whose efficiency depends on congestion avoiding and reliable transfer ability. Congestion control mechanism based on that packet losses due to buffer over flows. TCP is not suitable for wireless links since it cannot differentiate packet losses caused by congestion or link failure. Reduce transfer rate on observing packet losses even having no congestion and causes to lower the TCP's throughput over wireless link than actual [3] which is not justifiable. One solution to improve throughput in link failure case is not allow the TCP to reduce the transfer rate as it does in congestion case by hiding link failure from senders or retransmitting packets in data link layer [2] but require network support. Whereas the approach adopted here to impart a machine learning algorithm in TCP that classify causes of packet losses observing network behavior simulating by ns-2 etc which is suitable for machine learning techniques[1]. So proposed supervised learning algorithm built model for two causes of packet lose to enhance the performance of TCP over wireless and wired network.

Identification of network applications timely and accurately is growing need to observe traffic flows directly referred as classification [6]. Application identification is used for dynamic access control; Network based quality of Service, legal interception and firewalls. UDP or TCP port based classification is less effective due to increasing network applications allocating dynamic ports, non standard ports are used by user to hide traffic, and network address port translation is used.

Knowledge about payload formats required by payload based classification for every application i.e protocol decoding and signature matching requires knowledge of payload patterns. As classification rules updated due to protocol change so payload is inaccessible due to privacy laws and encryption are the limitations of this approach [8]. Traffic flow based application protocol independent inter-arrival times and packet length. For known network applications machine learning classifier is trained on set of flow instances and class of unknown flows are determined by this classifier to focus on achieving high accuracy. Many machine learning algorithms are used to achieve this high accuracy. Flow features used for IP traffic classification to reduce the consistency-based and correlation-based features reduction aspects of machine learning classification algorithms [5]. Computational performance improved by reducing the traffic flow features using feature reduction but it cannot reduce the classification accuracy. Same accuracy is achieved by machine learning algorithms Naïve Bayes Tree, C4.5, Bayes net etc.

The big problem of MANET routing protocols is that they relies too much on hop count calculation. Proactive protocols only calculate hop counts to find out the optimal path [7]. OLSR and TBRPF are examples of that kind of protocols. Reactive protocols such that AODV, only calculate hop counts to find out if the routing table entry should be updated or not, DSR checks hop counts for automatic route shorting [9]. Although calculating hop counts leads to per-node delay and varying link stability but still it is best way in routing decision making. Sometimes these hop count evaluating methods may lose their efficiency. Thus routing metrics evaluation is the more efficient and adaptable way of MANET routing protocol [10].

In this paper, we focus on proactive MANET routing protocols and integration of neural network and OLSR. MANET researchers have done a lot of work on reactive MANET routing protocols. Reactive routing works well when node mobility is high, whereas when traffic load and data communication among the nodes increases it becomes less efficient.

The routing metric which is per-node local queuing delay, is the major part of end-to-end delay which is so dynamic and it is determined by available bandwidth and traffic load. In order to increase the adaptability of the routing protocol, we must consider the queuing delay without imposition of QoS constraints on routing.

MANET proactive routing is improved by including predicting on queuing delays mechanism. Any form of measured routing metric values do not predict the network conditions due to the dynamics in a MANET. So we use neural network as the prediction method to deal with non stationary and the nonlinear dynamics of queuing delays. Many applications of neural network have been developed in the last twenty years, in industrial control and optimization, data clustering, ATM multiplexing and call admission control.

2. RELATED WORK

In [4] feature values packet length, flow duration inter-arrival time used to cluster flow by Expectation Maximization algorithm. To detect intrusion network flow running on ports, principal component analysis used in [5]. Based on greedy feature search in [6] network application identification is proposed to divide into separate clusters. In [7] applications map to quality of service classes using and linear discriminate analysis (LDA) and nearest neighbor (NN) to accurately separate traffic into classes by ML supervised algorithm. Application types differentiated usin Naïve Bayes supervised classifier in [3]. Useful features are selected by correlation-based feature selection to reduce feature set and classify accurately in [8] different application types classify traffic flow. Classification performed by five packets of flow in a k-means simple clustering algorithm used in [9].

Classification of causes of packet loss in wireless networks discussed in [1][4]. Considered two approaches, in first classification making rely on network support and needed Explicit Congestion Notification to classify loss due to congestion. In [5] information of wireless link loss rate is required proposed in this paper using a learning algorithm as classifier impart with TCP without changing error recovery and congestion control algorithm of TCP. By this technique congestion control algorithm ignor loss due to link failure and trigger transmit in normal way.

3. PROPOSED SUPERVISE LEARNING ALGORITHM

Based Model:

Automatic learning to classify packet losses in TCP is achieved by supervised learning based model to improve the performance of TCP over wireless and wired network. It predicts the behavior of network traffic accurately using decision tree and neural network.

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 2, Issue 3, pp: (22-25), Month: September-December 2015, Available at: www.noveltyjournals.com

This paper proposed an automatic classifier technique using supervised learning to differentiate losses in wireless and wired network's link error due to congestion. To build a model, gathered causes of packet loss using simulation tool and make a database used by supervise learning algorithm's decision tree and neural network to classify such losses due to congestion and due to link failure may happen again. On link error this imparting algorithm TCP prevents TCP from reducing window size to improve the throughput over wireless links when there is no congestion. The classifier has limitations needed to be focused in future work, first it cannot identify data packet losses due to timeout expiration, second the database for supervised learning algorithm built by simulation tool that can differ in real physical network environment but classifier is bias free since database included samples of usually maximum network conditions.

4. MACHINE LEARNING CONCEPTS

Network traffic flow instances map with network traffic classes by machine learning algorithms. Statistical feature set with values used to describe each flow. Standard deviation of Inter-arrival time, mean packet length used to calculate a feature which is a descriptive value. Each Traffic flow for same set of features reveals network traffic values. IP traffic classification by supervised machine learning algorithm must know the class of traffic flow before learning, in unsupervised based on same values traffic flow divided into clusters which are not pre defined. Classification model predict new instances of unknown flows' feature value, can be built by taking training set from each class.

Feature Reduction:

As feature is statistic value calculated from flow information, for both directions as network flow is bi-directional. Need to decide more useful features for IP classification since using maximum features to train a classifier can have negative impact on algorithm performance. Selecting useful features can be performed by machine learning feature selection algorithms.

Evaluation Techniques:

Classification model built by training set having different class network flows and unknown network traffic is classified by testing set having unknown examples from each class. Traffic flows in both sets are labeled so that classifier will predict flows. Algorithm perform well on unseen examples by dividing data set in k subsets, each time one subset for testing and k-1 for training sets. Classification accuracy measured by combining accuracy (% of correctly classified instance), precision (No. of correctly classified class members) and recall.

5. CONCLUSIONS AND FUTURE WORK

Traffic classification using payload based and port analysis gain interest to develop machine learning techniques focusing only on achieving accuracy using different datasets and features but effect of feature selection and performing computation on classification is not discussed. So to limit type and number of features traffic classifier would be introduced based on bayes network, C4.5, naïve bayes to calculate computational performance and classification accuracy by reducing feature sets having minimum impact on classification but differ in performing network traffic classification. For future work need to investigate that why these machine learning algorithms differ in classifying network traffic flow. In order to calculate the predicted delays in routing table evaluation, we merge the proactive MANET routing protocol and delay prediction mechanism. This paper proposed an automatic classifier technique using supervised learning to differentiate losses in wireless and wired network's link error due to congestion. To build a model, gathered causes of packet loss using simulation tool and make a database used by supervise learning algorithm's decision tree and neural network to classify such losses due to congestion and due to link failure may happen again. On link error this imparting algorithm TCP prevents TCP from reducing window size to improve the throughput over wireless links when there is no congestion.

6. FUTURE WORK

The classifier has limitations needed to be focused in future work, first it cannot identify data packet losses due to timeout expiration, second the database for supervised learning algorithm built by simulation tool that can differ in real physical network environment but classifier is bias free since database included samples of usually maximum network conditions.

ACKNOWLEDGEMENT

REFERENCES

[1] Ibtissam El Khayat, Pierre Geurts & Guy Leduc's "Enhancement of TCP over wired/wireless Networks with Packet Loss Classifier using Supervised Learning, " *Wireless Networks (10220038),* Vol.16 Issue 2, p273, Feb 2010.

[2] Nigel Williams, Sebastian Zander & Grenville Armitage's "A Preliminary Performance Comparison of Five Machine Learning Algorithms for Practical IP Traffic Flow Classification, " *ACM SIGCOMM Computer Communication Review,* Oct. 2006.

[3] Zhihao Guo & Behnam Malakooti's "Predicitive Delay Matric for OLSR using Neural Networks, " *Proceedings of the 3rd international conference wireless internet,* 2007

[4] Wei Lu, Mahbod Tavallaee & Ali A Ghorbani's "Hybrid traffic classification approach based on decision tree, " *IEEE conference* December 2009.

[5] Annie' Montigny Leboeuf, Mathieu Couture & Frederic Massicotte's "Traffic Behaviour Characterization Using NetMate, " *Springer*, September 2009.

[6] Panagis Magdalinos, Apostolos Kousaridas, Panagiotis Spapis, George Katsikas & Nancy Alonistioti's "Enhanceing a Fuzzy Logic Inference Engine through Machine Learning for a Self-Managed Network, " *Mobile Network and Applications,* Vol.16, Issue 4, p 475-489, August 2011.

[7] Barman D & Matta's "Effectiveness of loss labeling in improving TCP performance in wired/wireless networks, " *IEEE*, 2002.

[8] Yong Wang, Margaret Martonosi & Li Shiuan Peh's "A Supervised Learning Approach for Routing Optimizations in Wireless Sensor Networks, " *o6 proceedings of the2nd international workshop on multi-hop ad hoc networks ACM,* May 2006.

[9] Zhihao Guo, Shaya Sheikh, Camelia Al Najjar & Hyun Kim & Behnam Malakooti's "Mobile ad-hoc network proactive routing with delay prediction, " *Wireless Networks ACM,* Vol. 16, Issue 6, p 1601-1620, October 2009.

[10] Ddd A. W. Moore, D. Zuev's "Internet Traffic Classification using Bayesian analysis Techniques, " *SIGMETRICS 05 Proceedings of the 2005 ACM SIGMETRICS international conference on measurement and modeling of computer systems ACM*, June 2005.

[11] S Zanader, T. Nguyen & G Armitage's "Automated Traffic Classification and application Identification using Machine Learning, " *Local Computer Networks IEEE,* November 2005.