ORIGINAL RESEARCH PAPER

Computer Science

NETWORK CONGESTION PREDICTION USING CLASSIFICATION

KEY WORDS: Network Congestion, Hacking, Scalability, Congestion detection First Section

Dr A Prasanth Rao	Department Of It, Anurag University, Hyderabad
Akula Venkata Sai Akhil	Department Of It, Anurag University, Hyderabad
Akku Vivek	Department Of It, Anurag University, Hyderabad
Preetham Reddy Yaramada	Department Of It, Anurag University, Hyderabad

ABSTRACT

Network congestion is a situation that happens when a network is un- able to handle data is more than its threshold value when load becomes high. This situation is known to be network congestion which causes packets to be dropped on the network due to overflow of buffer and therefore leads to data loss and unreliable connection. Therefore, effective congestion control is an important is- sue that needs to be addressed in the transport layer. There are various factors such as hardware, software and miscellaneous factors can lead to network con- gestion. Hardware factors include non-compatible hardware, outdated routers and too many devices connecting to a single router. Software factors include some of the devices following their own protocol, ineffective communication protocols, improper fire walls and proper limitations not defined on the file sizes to be transferred. Poor network design, network hacking and over subscription leads to the miscellaneous factors. As a consequence of such factors, network performance will degrade dramatically, and system performance will be affected. This is an undesirable condition that needs to be corrected. Thus, our model uses classification algorithms which help in predicting network congestion before- hand, thuspreventing the packet loss and damage.

1. Introduction

The network congestion is one of the research areas due to rapid increase in scalability of network devices. There are many reasons network is congested that de- grade the performance and provide reduced quality of service. The network congestion happens when the communication channel cannot handle huge amount of data that con- nects two or more devices which higher than its threshold value. This congestion prob- lem reduces throughput and performance due to packet loss, delay in queuing packets and blockage of certain connections. The hardware, software and miscellaneous are the main features of network congestion and our goal asses network congestion prediction is done before it has occurred. The congestion data set (csv file) consist of all features like hardware, software and miscellaneous inputs as column elements. The target label which provides three output labels (No, Low, High) of congestion values which is rep- resented as congestion problem in the network. The two algorithms AdaBoost and SVM to carry out the classification process and help to learn the system based on the predic- tions. This paper aims at analyzing a data set that contains several hardware, software and miscellaneous factors that may cause congestion. Each factor has 2 values associ- ated with it that determines the extent of the congestion occurrence.

Priority Based Congestion Control Method used packet service time as a metric for congestion in which transmission rate is set by using upward node transmission rate and packet priority [1]. Campbell proposed a method called Congestion Detection and Avoidance that makes use of wireless channel load and queue length as one of the met- rics for congestion [2]. The rate of transmission is reduced when the source does not receive the acknowledgement for the packet transferred. The authors S. Deering and R. Hinden use a method called CONSEQ which is a hop-by-hop congestion control and load balancing scheme. This method uses Effective Queue Length (EQL) as the con- gestion metric along with fuzzy logic when congestion is not controlled In order to improve the network throughput and lifetime [3]. A. Al-Kashoash, H. M. Amer, L. Mihaylova, and A. H. Kemp propose a protocol which uses 2 strategies - traffic control and resource control strategy called an Optimization based Hybrid Congestion Control. This uses multi-attribute optimization methodology so as to forward the packets in a non-congested path to achieve throughput and efficiency [4]. The authors C.-Y.Wan, S. B. Eisenman, A. T. Campbell, and J. Crowcroft propose a method called DDR wherein buffer occupancy is higher than maximum threshold. The new transmission rate is then set by using queuing theory and the least hop count is measured and then assigned to the node which is closer to the sink node.

Organization of this paper as follows- Section 2 presents Literature Survey, Sec- tion3 Architecture Diagram for network congestion prediction Algorithm Section 4 pre- sents Results and discussion Section 5 presents conclusion and future scope.

2. Architecture Diagram for Network Congestion Prediction

An architecture diagram gives detailed description about the structure of network con- gestion structure, organization and its behavior. This model is conceptual model that defines all elements required to analyze the problems and its solutions to predict the Network congestion. Network Congestion System mainly consists of database with several entities are attached to it. The data plot generated from data sets which represent the division of data rows and these data are used input for our proposed algorithms.

The two algorithms such as AdaBoost and SVM are used to predict and meas- ure the performance of the system. The database consists of several features such as Hardware, Software and miscellaneous used in network and those values helps us to analyze behavior of congestion problem. The next section presents the algorithm to predict and accuracy of the system.

Working Process Model

The working process model presents in the dia- gram Figure 2.1 which is used to detect congestion problem. The network consists of multiple nodes and each node in the network can be evaluated where congestion is to be detected on basis of types of packet statistics. The packet in each node and number of packet (Sent, received) can be evaluated using output label (No, Low, High) values. Once feature extraction process completes then feature collection process imitates which collects the data for measuring congestion problems.

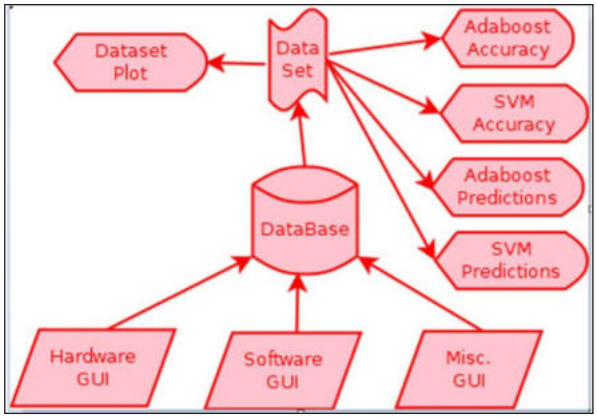


Figure 2.1: Network Congestion Architecture Diagram

Algorithm of Adaboost

The main purpose of this algorithm is to construct a decision tree for each feature. Initially, a stump is created for features such as hardware, software and miscellaneous and it is known to first base learner model. The performance of the system can be com- puted using incorrectly assigned weights to the records and correspondingly new mod- els are generated for further classification. This process repeated till upgraded dataset is created in which a normalized weights are allocated to each record and system learns effectively to predict network congestion.

Step 1:

A decision tree constructed on the top of training data using weighted samples.

The weight of each sample indicates the importance of each sample and also its classi- fication.Intially all samples given equal weights.

Step 2:

For each feature, a decision tree is created which gives information about

samples in their target classes.

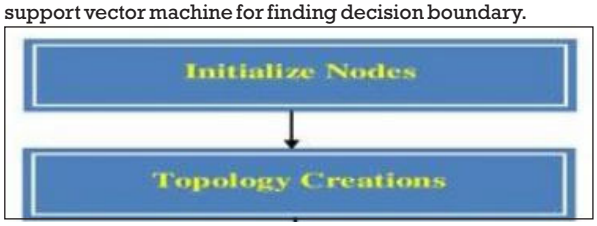
Step 3:

For each classification, weight assigned and each weight is dynamically con- figured to measure accuracy of the classifier.This means high accuracy = high weight.

Step 4:

Repeat from Step 2 until all the data points have been correctly classified.

There is need to combine all such features together using support vector machine for finding decision boundary.



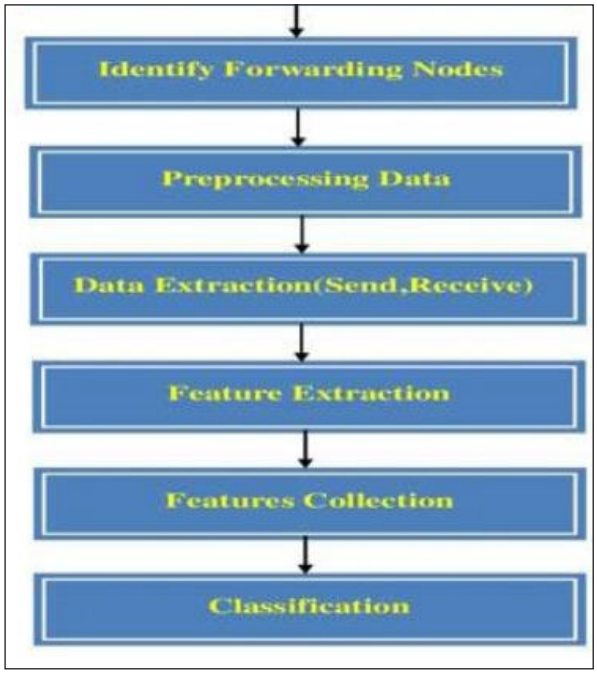


Figure 2 .2Netwotk Mode

The Support Vector Machine algorithm intends always to find a hyper- plane in an N- dimensional space. An N dimensional space refers to a space with N number of features. This hyper plane should be able to distinctly classify all the data points. The separation of data points can be done using many hyper planes. Several hyper planes can be found, but our motive is to find a plane which will have the highest margin, i.e. it should have the maximum distance between the data points of both the classes. Such a margin is desirable. This is called the maximum marginal hyper plane. The closest of all data points is known to be decision boundary.

3. Simulation Results

The simulation results generate using python programming Language using libraries and function for predicting network congestion. The output pie chart of input data figure 5.1 shows that 20% testing data where as 80% Training data of total data sets.

Output labeled Data:

Distribution of dataset Total rows = 3600

Training: 2880 rows are taken as training data (2880/3600)*100 = 80% Testing: 720 rows are taken as testing data (720/3600)*100 = 20%

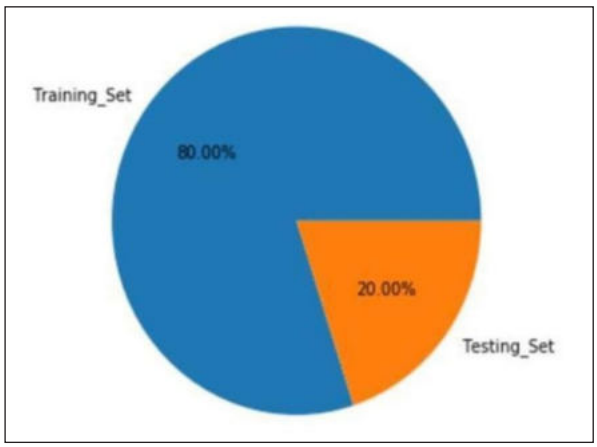


Figure 3.1: Output Pie Chart of input data

Performance Calculation using Adaboost Classifier

The accuracy of AdaBoost classifier is 80.38% [1] - No Congestion [2] –Low Con- gestion [3] – High Congestion

```
The accuracy of ABClassifier on testing data is: 80.3894297635605
ABClassifier Prediction on the first test set is: [2]
ABClassifier Prediction on the second test set is: [2]
ABClassifier Prediction on the third test set is: [3]
```

Figure 3.2 Adaboost Classifier SVM Classifier Performance

The accuracy of SVM classifier is 96.94% [1] - No Congestion [2] – Low Conges-

tion [3] – High Congestion

```
The accuracy of SVM Classifier on testing data is: 96.94019471488178
SVM Prediction on the first test set is: [1]
SVM Prediction on the second test set is: [2]
SVM Prediction on the third test set is: [3]
```

Figure 3.3: SVM Classifier

4. Result Analysis

The performance of two congestion prediction algorithms such as Adaboost and SVM implemented and tested. Our Observation shows that the SVM gives a better ac- curacy as compared with Adaboost and SVM gives an accuracy of 96.54% whereas Adaboost gives an accuracy of 80.38% . Hence it's concluded that our model works best is used for prediction of network congestion.

5. Conclusion

This project titled ‘‘Network congestion Prediction using Classification’’ is useful in understanding about the factors leading to network congestion. The project is useful to the data analysts to understand more about the Adaboost and SVM classifications and the role that factors play in altering the performance of the algorithms.This project finally leads to the improvement of network services