ISSN No:-2456-2165

Loss Classification Analysis in Wireless Networks

Mohammad Ummer Chopan
M-tech Scholar CSE
Sharda University
Greater Noida, India

Pooja
Associate Professor, CSE
Sharda University
Greater Noida, India

Amit Upadahay
Assistant Professor, CSE
Sharda University
Greater Noida, India

Abstract:- Congestion control mechanism in wireless networks in vogue has been studied extensively over the past decade with the sole aim of improving the performance over the wireless links. With the extensive increase in the use of wireless technology, the idea of improving the TCP performance has widely gained currency. In this paper we are going to present an approach of loss classification for enhancing the TCP performance in wireless networks.

Keywords:- *Loss Classification, Machine Learning, TCP, Wireless Networks.*

I. INTRODUCTION

TCP carries 90% of the internet traffic making it the most widely used and reliable transport protocol. One part of TCP that could utilize improvement is it execution over remote connections. When TCP detects a lost segment it reacts by resetting the congestion window and cutting the slow start threshold. The initial objective of TCP was to effectively use the available bandwidth in the network and to abstain from over-burdening the system by properly throttling the sender's transmission rates[1]. Network congestion is deemed to be the underlying reason for the packet loss and is the only loss cause identified by the existing TCP. Consequently TCP performance is often unsatisfactory when used in wireless networks and requires various improvement techniques or methods to overcome this problem of labelling all losses to a single cause. An important factor causing the unsatisfactory performance is that TCP is unable to differentiate losses which occur due to congestion from those which occur due to link errors, which is more likely the case in wireless networks[2].

A. TCP CONGESTION CONTROL

TCP congestion control mechanism in vogue is based on the concept of congestion window and the size of this congestion window, which is adapted when necessary according to a required procedure.

Congestion usually arises in the network when routers are unable to process the data that arrive at them which is mainly termed as buffer overflow in networking. This results in the packet loss. The mechanism TCP employs works by increasing the rate steadily (additive increase) when in the event a loss is perceived, reduce it more abruptly (multiplicatively) as soon as loss occurs.

In TCP usual mechanism involves sending an acknowledgment by a receiver in response to a packet from the sender. The sender keeps sending packets as it receives the acknowledgment for every packet sent. When it gets the acknowledgment of the last packet of the previous group of CWND, It increments CWND by 1 and then sends two packets in burst instead of one. The receiver realises that a packet has been lost when it receives a packet before its predecessors in the sequence order. In this situation, it sends to the sender , the acknowledgment of the last packet it has received in the sequence as packets may not be received in order for some other reasons than packet loss, since this is the only reason entertained in such cases. A packet loss is detected simply after three duplicate acknowledgments. As soon as the sender receives the three duplicate acknowledgments it divides the congestion window multiplcatively ,reduces the data traffic flowing through the channel and consequently starts retransmission. [4].

The existing congestion control mechanism attributes all losses to buffer overflows and therefore reduces the traffic rate as soon as it senses congestion. This mechanism works well in wired networks but when it comes to wireless networks this leads to massive throughput problems because the losses that mainly occur in wireless networks are due to link errors and not due to congestion and the existing TCP has no such mechanism of distinguishing these two different loss causes.

II. PROPOSED SOLUTIONS

A straightforward solution is to prevent the TCP from reducing the traffic when it faces packet loss due to link error as it does when it occurs due to congestion. This solution can be achieved by using machine learning approach as well as fuzzy logic. Both these techniques can be of great use in classifying losses.

A. MACHINE LEARNING APPROACH

Machine learning involves the process of training machines by feeding them data, it allows machine to automatically learn from the data without be explicitly programmed This uses information attainable at the transport layer [5]. By using various machine learning algorithms, the loss causes can be easily classified and thereby enabling the TCP to react only to those caused due to congestion. In machine learning supervised learning techniques are generally chosen for classification problems. A supervised learning algorithm receives a learning sample and returns a function f(an hypotheses or model) which is chosen in a set of candidate funcitions. Usually, the sample of observations is called the learning sample LS and is a set of input/output pairs .

LS={x1/y1,,..xn/yn} where xi is the input vector value(additionally called the attributes) relating to the ith observation (additionally called an object) and yi is its output vector value. Attribute values may be discrete or continuous. At the point when the yield or output takes its qualities in a discrete set say {C1, C2,.. , Cm}, we talk about a classification problem and when it is continuous, we talk about a regression problem. The fundamental standard employed to evaluate learning algorithms is their ability to accurately make predictions i.e. the manner in which the model they deliver sums up to unknown/inconspicuous data. More often than not, the ranking among different algorithms depends to a great extent on the issue and how well the essential theories of the learning algorithms are fulfilled by this issue. Another essential factor is interpretability followed by the computational efficiency of the learning algorithm. By and large, there is a trade-off between these three criteria.

B. SOME MACHINE MACHINE LEARNING TECHNIQUES

- *K-nearest neighbour*

A machine learning algorithm used for classification purposes[6]. This is the simplest technique . It stores all available cases and classifies new cases based on similarity measure (distance function). An important drawback of this technique for this application is that the calculation of prediction is quite demanding.

- *Logistic Regression[7]*

Logistic regression is a statistical method for analysing the dataset in which there one or more independent variables that determine an outcome. The outcome is measured with a dichotomus variable I;e it contains data coded as 0 or 1.

- *Decision trees[8]*

Decision tress are often used for classification purposes due to its being simple and easily understandable. It is a flow- chart like tree structure where each internal node represents a test on an attribute, each branch represents an output of the test and class label is represented by each leaf node.

C. FUZZY APPROACH

- *FUZZY Approach*

While traditional logic contains only two truth values (true and false or 1 or 0),fuzzy logic may contain an infinite number of truth values on the continuous range (0,1). A Fuzzy set is a set whose memberships are defined by a function that maps the elements of the set to [0, 1].The mapping functions are called membership functions and there may be an arbitrary number of them on any domain[9].

Fuzzy logic and fuzzy inference can be effectively used for classifying loss causes. Simple fuzzy approach has been discussed in several but we are proposing a different approach such as fuzzy inference engine which is composed of a set of input and output variables. A set of linguistic rules defined inside of an inference engine maps the inputs to the output. The rules follow the modes ponens (IF THEN) form and an inference engine may have an arbitrary number of them defined.

Once a membership function has been defined a fuzzy rule base must then be defined. A fuzzy rule base is simply a series of implications consisting of a variable number of antecedents and consequents.These rules are defined in terms of linguistic variables and values. Inputs to the system can be crisp values that take the form of delta functions after fuzzification.[10] The minimum value of inputs is intersected with the membership function. For every rule a derived fuzzy set is produced and the results aggregated. This process produces the final inference result which is a fuzzy set with low membership values for small values of the parameter , large membership values for medium values parameter and no membership values parameter used. Because the result is a fuzzy set therefore a technique called defuzzification is used for transforming a fuzzy set into a crisp value. There are various defuzzification methods and one such method is a centroid method which can be used for the sake of brevity. The centroid method is defined as follows:

$$y^* = \frac{\int \mu_c(y) \cdot y \, dy}{\mu_c(y) \cdot dy}$$

Where Y* is a scalar representing the crisp inference y is the domain of the output is the implied membership of the rule base.

The inference system defines two linguistic terms{increasing , stable} for describing the variances of the inter arrival times because variances are good indicators that there is some atypical congestion on the network . It stands to reason that a drop to a link error should have little correlation to the variances experienced prior to that drop. However a drop due to congestion may have indeed be foreshadowed by fluctuating inter arrival times. To calculate variance we store the last two round trip times worth of inter arrival times[11]. When a loss occurs the ratio for the variances of the two round trip times are compared against each other. A ratio less than one suggests that variance has increased in the last couple of a RTT's. RTT or round-trip-time is the total amount of time between the sending of packet and the receiving the acknowledgment.And intuitively an increase in variance suggests congestion in a network. The inter arrival input is then defined as the variance of the second to last RTT. If the variance is decreasing we can only assume that the network is stabilizing, perhaps indicating that congestion is not the root cause. In some situations where congestion is experienced, the delays experienced on one side of the network will not be equal to the delays going in the other way. We use this intuition as a metric for gauging loss causes .The second input is then defined as the ratio of the measured delay to the estimated RTT[12]. The closer the ratio to one the more stable the network is, which equates to a lower probability of congestion. Linguistics variables for describing one way delay{High, medium, low}.Now if a loss event occurs due to congestion, we would expect a high

ISSN No:-2456-2165

variance in the inter arrival times and high one way delays. Table 1 represents the entire fuzzy rule base.

Inter arrival variance	One way delay	Loss cause
Increasing	High	C
Increasing	Medium	C
Increasing	Low	C
Stable	High	C
Stable	Medium	LE
Stable	Low	LE

Table 1:- Showing fuzzy rules

III. PERFORMANCE EVALUATION

A standout amongst the most widely recognized and suitable technique for assessment is a confusion matrix shown in table 2 and is usually formed from the four outcomes. These outcomes are mainly used the indicators of the classifier performance and the assessment of these values help determine the accuracy and other notable parameters. These four outcomes as a result of a binary classification are:

- True Positive (TP) which is the correct positive prediction.
- False Positive (FP) which is the incorrect positive prediction.
- True Negative (TN) which is correct negative prediction.
- False Negative (FN) which is the incorrect negative prediction.

		Predicted	
		Positive	***Negative***
Observed	***Positive***	TP	FN
	Negative	FP	TN

Table 2:- The Confusion Matrix Of Two Class Classification Problem.

Along with accuracy certain other measures such as precision, recall and F score can also be derived through the values obtained from the confusion matrix. The evaluation of these addition measures gives a better look at the performance metric.

IV. CONCLUSION

Here in this paper, We have presented certain approaches that can be invoked for classifying loss causes. We centred our approach around supervised learning and around fuzzy inference in fuzzy logic. Pragmatically both approaches are helpful in classifying losses consequently help in the improvement in the performance over wireless links.

REFERENCES

[1]. “Balakrishnan, H., Padmanabhan, V. N., Seshan, S., & Katz, R. H.,“A comparison of mechanisms for improving TCP performance over wireless links,” In Conference Proceedings on Applications, Technologies, Architectures, and Protocols for Computer Communications pp. 256–269), 1996, ACM Press.”

[2]. “Xylomenos, G., Polyzos, G., Mahonen, P., & Saaranen, M.,“TCP performance issues over wireless links,” Communications Magazine, IEEE, 39(4), 52–58, 2001.”

[3]. “Wu, H., Peng, Y., Long, K., Cheng, S., & Ma, J.,“Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and enhancement,” In Proceedings of INFOCOM 2002, Joint Conference of the IEEE Computer and Communications Societies (Vol. 2, pp. 599–607), 2002.”

[4]. “Hala ElAarag, “Improving TCP Performance over Mobile Networks,” Journal of ACM Computing Surveys , Vol. 34, No. 3, pp.357-374, September 2002.”

[5]. “Biaz, S., & Vaidya, N. H.,“ Distinguishing congestion losses from wireless transmission losses,”: A negative result. In Proceedings of IC3N, New Orleans, 1998.”

[6]. “Barman, D., & Matta, I. ,“ Effectiveness of loss labeling in improving tcp performance in wired/wireless networks,” In Proceedings of the 10th IEEE International Conference on Network Protocols (pp. 2–11),IEEE Computer Society, 2002.”

[7]. “Liu, J., Matta, I., & Crovella, M. “End-to-end inference of loss nature in a hybrid wired/wireless environment.,” In Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2003.”

[8]. “Liu, J., Matta, I., & Crovella, M. “End-to-end inference of loss nature in a hybrid wired/wireless environment.,” In Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2003”.

[9]. “Parsa, C., & Garcia-Luna-Aceves, J. J. , “Differentiating congestion vs. random loss: A method for improving TCP performance over wireless links,” In Proceedings of IEEE WCNC’2000 (pp. 90–93), 2000.”

[10]. “ElAarag H. and Bassiouni M.,"Performance Evaluation of TCP Connections in Ideal and Non-Ideal Network Environments,” Computer Communications Journal, ElSevier Publications, v 24 n 18 pp. 1769-1779, Dec 1 2001.”

[11]. “Hala ElAlarag, Matt Wozniak, “Using fuzzy inference to improve congestion control in wireless networks ,” 2012.”

[12]. “Shi, Kai, Yantai Shu, and Qingfeng Song,"Receiver Centric Fuzzy Logic Congestion Control for TCP Throughput Improvement over Wireless Networks," Proc. of International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China. Web. 16 Apr. 2011.”