# Cluster Based Packet Loss Prediction using TCP ACK packets in Wireless Network

BALAJI V
Research Scholar,
Hindustan College of Engineering & Technology
Coimbatore,, India

DURAISAMY V
Principal
Hindustan College of Engineering & Technology
Coimbatore, India

***Abstract*— With the rapid growth of wireless networks, the wireless portion of the network becomes the bottle neck and a major source of congestion. Planning in the area of wireless network becomes crucial for a sturdy network providing expected quality of service. This paper studies the correlation of the tcp ack flag with respect to network congestion. We describe a clustering approaches to determine the status of access point for a given channel. This will help to determine and plan for additional channel allotment during peak loads. We have used the traces collected by wireless monitoring at the 62nd Internet Engineering Task Force (IETF) meeting held in Minneapolis, MN, March, 2005.**

**Keywords- Congestion, Tcp flags, Wireless monitoring, Clustering, Access point.**

## I. INTRODUCTION

Wireless networks have been extensively studied in recent years by a growing number of researchers. It is desirable that the wireless network loan is anticipated for any given access point. Traditional resource management passively tracks the network and collects the wireless data. Existing channel allocation and borrowing algorithms have been investigated and suggestions made. The goal for any service provider is towards optimal wireless network resource management. This affects all aspects the organization including capacity planning, operations and allocation to meet the increasing demand on planning management and operations of the wireless network. The operation and planning of a wireless network requires an appropriate model for predicting the mobile traffic load.

An wireless network with a high density of nodes and within a single collision domain has a high probability of congestion, decreasing the performance significantly. Typically congestion include drastic drops in network throughput, unacceptable packet delays and session disruptions. Typically the back-haul wireline portion of a wireless network is typically well provisioned to handle the network load. Therefore, there arises a compelling need to understand the behavior of the wireless portion of heavily utilized and congested wireless networks[1].
The evaluation of wireless network requires the generation of workloads to test the viability and performance of the new protocol or technique being studied. Lack of realistic traffic is a major limitation that has forced researchers to generate synthetic data for their simulations and experiments to evaluate performance of wireless technologies. Uniform constant bit-rate traffic (CBR) is typically used to evaluate networking protocols, but there is no evidence that behavior under such workloads is an accurate predictor of performance under real usage patterns. The inability to experimentally forecast real-world performance is a severe handicap to the wireless networking community. It hinders the ability to effectively develop better solutions to the many difficult problems that face emerging wireless technologies[2].

## II. MOTIVATION

A number of studies of scaled up wireless network deployments have been carried out by M. Balazinska et al [3], D. Kotz et al[4], T. Henderson et al[5] and A. Jardosh et al[6]. These studies analyze a wide range of wireless network behavior and provide insights into the behavior of deployed networks in different scenarios. Most of the studies have been conducted within the wireless LAN of university campuses. In the process of their research huge amounts of raw wireless traffic data has been collected for subsequent analysis and research. In this work we propose to use such raw data and use this data to create a model for congestion prediction. The reason for choosing real time data is that it is not only limited to simulations but also can be used for experimental deployments which is a necessity for wireless protocol evaluation.

Congestion has been considered as one of the basic issues for both wired and wireless networks. The ultimate goal of congestion control remains very same for also type of communication networks. Bandwidth, traffic rate and data transfer rate differ for different types of network. Some protocols like ATM handle busty data transfer with their own congestion control and avoidance schemes. Optical network handles congestion with optical packet buffering. Typically congestion control mechanism becomes a key feature for QOS [7]. Initially congestion control started with reactive techniques. Now predictive techniques are becoming popular to the problem of congestion. While there are a multitude of protocols in use in the Internet, the most prevalent is the Transmission Control Protocol (TCP).To avoid congestion collapse, TCP uses a multi-faceted congestion control strategy.
For each connection, TCP maintains a congestion window, limiting the total number of unacknowledged packets that may

be in transit end-to-end. This is somewhat analogous to TCP's sliding window used for flow control. TCP uses a mechanism called slow start to increase the congestion window after a connection is initialized and after a timeout. It starts with a window of two times the maximum segment size (MSS). Although the initial rate is low, the rate of increase is very rapid: for every packet acknowledged, the congestion window increases by 1 MSS so that the congestion window effectively doubles for every round trip time (RTT). When the congestion window exceeds a threshold the built in algorithm enters a new state, called congestion avoidance[8]. However in a wireless network temporary losses due to fading, shadowing, and other radio effects, that cannot be considered congestion also occur which needs to be taken into consideration.

## III. GOAL OF OUR WORK

The objective of this paper is to investigate raw wireless trace files and the development of a clustering model, Packet Loss Based Cluster Prediction (PLBCP) suitable for the analysis of TCP Ack flags for packet loss prediction.

It is known that some TCP segments carry data while others are simple acknowledgements for previously received data.The popular 3-way handshake  tilizes the SYNs and ACKs available in the TCP to help complete theconnection before data is transferred. Each TCP segment has a purpose, and this is determined with the help of the TCP flag options, allowing the sender or receiver to specify which flags should be used so the segment is handled correctly by the other end. The ACKnowledgement flag is used to acknowledge the successful receipt of packets. The foundation of our work is based on the packet retransmission detection through the detection of duplicate ack packets in the trace file.

For our analysis, we use data from a wireless network deployed at Internet Engineering Task Force (IETF) meeting held in Minneapolis on March 2005. The IETF network was comprised of 38 Airespace2 1250 access points (Aps) distributed on three adjacent floors. Each Airespace AP supported four virtual Aps. A virtual AP is a logical AP that exists within a physical device and enables the wireless LAN to be segmented into multiple broadcast domains. This provides the ability to map multiple Extended Service Set Identifiers (ESSIDs) to multiple Basic Service Set Identifiers (BSSIDs). Thus, at the IETF, a total of 112 Aps (38 physical Aps x 4 ESSIDs per physical AP) were available for utilization[8]. We limit our work to over a 30 minute interval and consider only the TCP and UDP packets.

Figure 1 shows the TCP and UDP traffic flow captured during 30 minutes.

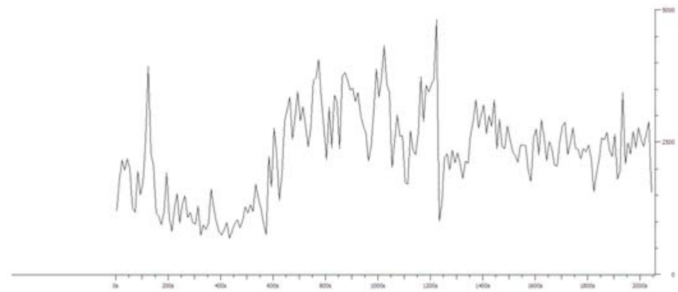
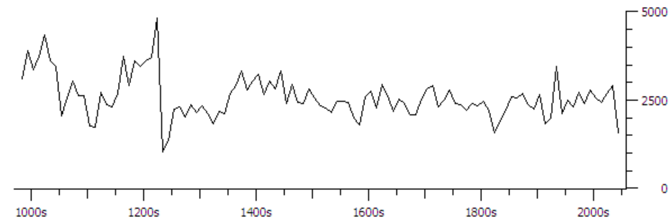


Figure 2. shows the duplicate acknowledgement packets received due to packet loss.




The Multi Step cluster method we propse is a scalable cluster analysis algorithm designed to handle very large data sets. In the first pass the data is pre clustered to smaller clusters. In the second pass the smaller clusters are broken down to still smaller clusters. The pre-cluster step uses a sequential clustering approach. It scans the data records one by one and decides if the current record should be merged with the previously formed clusters or starts a new cluster based on the distance criterion mentioned herein. The procedure is implemented by constructing a modified cluster feature (CF) tree. The CF tree consists of levels of nodes, and each node contains a number of entries. A leaf entry (an entry in the leaf node) represents a final sub-cluster. The non-leaf nodes and their entries are used to guide a new record quickly into a correct leaf node. If the CF tree grows beyond allowed maximum size, the CF tree is rebuilt based on the existing CF tree by increasing the threshold distance criterion. The rebuilt CF tree is smaller and hence has space for new input records.

A log based distance measure can be used to calculate the distance between clusters. The distance between two clusters is related to the decrease in log likelihood as they are combined into one cluster. The distance between clusters j and s is defined as:

$$d(i,j) = \xi_i + \xi_j - \xi_{<i,j>}$$

where

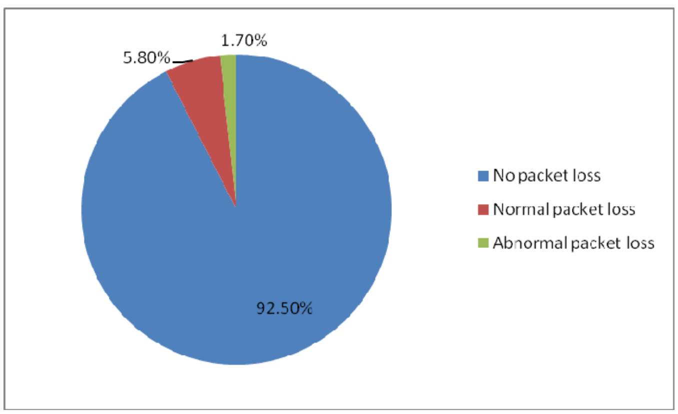
$$\xi_v = -N_v \left( \sum_{k=1}^{K^A} \frac{1}{2} \log\left( \hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2 \right) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right)$$

and

$$\hat{E}_{vk} = -\sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v}$$

The clustering result so obtained is given below. From the pie chart it is seen that over 1.7% of the packets had abnormal losses resulting in re transmission.

Figure 3:The Distribution of Packets Discovered by the multi step cluster




## IV CONCLUSION

We proposed a multi-step clustering algorithm for clustering out the time periods where an abnormally high packet loss has been discovered due to either congestion or other effects common in wireless networks. The clustering algorithm proposed was able to cluster the data with very good efficiency. Further studies need to be done for a longer time frame.

## REFERENCES

[1] Amit P Jardosh, Krishna N Ramachandran, Kevin C Almeroth, Elizabeth M Belding-Royer "Understanding Congestion in IEEE 802.11b Wireless Networks" University of California.

[2] Stefan Karpinski, Elizabeth M. Belding, Kevin C. Almeroth "Towards Realistic Models of Wireless Workload", University of California.

[3] M. Balazinska and P. Castro, "Characterizing mobility and Network usage in a corporate wireless local-area network," in ACM MobiSys, San Francisco, CA, USA, May 2003, pp. 303–316.

[4] D. Kotz and K. Essien, "Analysis of a campus-wide wireless network," in ACM MobiCom, September 2002.

[5] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in ACM MobiCom, September 2004.

[6] A. Jardosh, K. Ramachandran, K. Almeroth, and E. Belding-Royer, "Understanding link-layer behavior in highly congested IEEE 802.11b wireless networks," in ACM Sigcomm EWIND, Philadelphia, PA, USA, August 2005.

[7] http://hubpages.com/hub/congestion-control

[8] http://en.wikipedia.org/wiki/TCP_congestion_avoidance_algorithm