Classification of application based on multiple cluster models with selected attributes from Flow statistics

Anil Kumar
Department of
Computer Engineering
Texas A
Commerce, Texas
Email:akatta1@leomail.tamuc.edu

M Commerce

Dr.Jinoh Kim
Twentieth Century Fox
Commerce, Texas
Email: homer@thesimpsons.com

James Kirk
and Montgomery Scott
Starfleet Academy
San Francisco, California 96678-2391
Telephone: (800) 555–1212
Fax: (888) 555–1212

Abstract—Identifying applications are critical for a broad range of network related activities like bandwidth usage, security etc. Earlier, applications are identified based on port numbers, which proved to be not accurate anymore; based of payload signatures, which is proved to be accurate but has been limited in the real world implementation because of privacy concerns; based on flow statistics, which uses machine learning algorithms to find the patterns in the flow statistics and use it in classification, which has been widely used for many classification problems. In this research, we explore the importance of the attributes or a combination of flow attributes which can classify applications effectively. The idea is to combine clustering and using combinations of flow attributes and we measure accuracy of each combination. We are currently evaluating our model with real-world traffic traces indicating effectiveness of the selective attributes is effective than using the whole set of attributes.

I. INTRODUCTION

Recently, there has been a lot of emphasis on using flow statistics with combination of Machine learning algorithms to determine network based applications. Accurately identifying network based application is of major interest for ISP. Which lets providing Quality based Service(limiting usage of bandwidth by unnecessary applications).

In this paper, we proposed a new technique which studies the relevance of flow statistics or combination of them to give us better classification. We compared our technique with previously published work. As with most machine learning algorithms, more attributes(not very important)[?] may actually hurt the overall accuracy of the classification. So, determining the best combination of attributes and also combination of them helps in actually improving overall accuracy. We developed our own technique, where we used multiple trained models to classifying the incoming flow.

II. RELATED WORK

Subsection text here.

III. IMPORTANCE OF ATTRIBUTES IN CLASSIFICATION

We have considered a total of 18 flow statistics in our studies.

Average IAT	Maximum of IAT	Min
Stadard Deviation of IAT	Average RIAT	Maxi
Minimum of RIAT	Standard Deviation of RIAT	Avera
Minimum of Packet Size	Maximum of Packet Size	Standard De
Total Packet Size	Flow Duration	Num
Average Packets/second	Average Bytes/second	Pa

IV. CONCLUSION

The conclusion goes here.

ACKNOWLEDGMENT