Abstract

Now a days internet traffic classification is an emerging research field since 1990’s because of its use in a large number of network activities. Traditional techniques of internet traffic classification that relied on well known TCP/UDP port numbers or payload based are rarely used because of use of dynamic port numbers instead of fixed port numbers and due to various cryptographic techniques which inhibit inspection of packet payload. Recent trends are use of ML (machine learning) algorithms for internet traffic classification. In our research work we propose a technique to classify the internet traffic into two classes, one for educational websites
and another for non-educational websites. In educational institutes for the optimum use of
network resources and for the welfare of the students, the use of non-educational websites
should be banned while only the educational websites should be allowed to open. To classify
the internet traffic we propose a technique to capture data packets first, related with various
educational and non-educational websites, using a packet capturing tool Wireshark. Then using
feature selection algorithm, a reduced feature dataset will be developed. After that training and
testing of various ML algorithms will have to be performed. Finally comparative analysis of the
different classifiers from the obtained results is to be performed.

Reference

1. Arthur Callado, Carlos Kamienski, Géza Szabó, Balázs Péter Ger´o, Judith Kelner, Stênio
Classification using Machine Learning,” IEEE Communications Survey & tutorials, vol. 10, no. 4,
3. Runyuan Sun, Bo Yang, Lizhi Peng, Zhenxiang Chen, Lei Zhang, and Shan Jing, “Traffic
Classification Using Probabilistic Neural Network,” in Sixth International Conference on Natural
5. Andrew W. Moore, Denis Zuev, Michael L. Crogan, “Discriminators for use in flow-based
classification,” Queen Mary University of London, Department of Computer Science, RR-05-13,
August 2005.
6. Hyunchul Kim, kc claffy, Marina Fomenkov, Dhiman Barman, Michalis Faloutsos, and
KiYoung Lee, “Internet Traffic Classification Demystified: Myths, Caveats, and the Best
7. A. Madhukar and C. Williamson. “A Longitudinal Study of P2P Traffic Classification.” In
MASCOTS’06, Monterey, USA, August 2006.
8. A.W.Moore and D.papagiannaki, “Toward the accurate Identification of network
applications”, in poc. 6th passive active measurement. Workshop (PAM), mar 2005, Vol.3431,
pp 41-54.
IMC’03, Miami Beach, USA, October 2003.
Application Identification using Machine Learning”. In LCN’05, Sydney, Australia, November
2005.
on Statistical Flow Characteristics". In PAM’05, Boston, USA, March 2005.
flow-based network traffic classification: Evaluation and comparison,” Performance Evaluation
A Proposal for IP Traffic Classifier for Educational Institutions


Index Terms

Computer Science

Communications

Key words

Machine Learning

Features