Accepted Manuscript

Imbalanced Traffic Identification Using an Imbalanced Data Gravitation-based Classification Model

Lizhi Peng, Haibo Zhang, Yuehui Chen, Bo Yang

PII: S0140-3664(16)30216-X
DOI: 10.1016/j.comcom.2016.05.010
Reference: COMCOM 5331

To appear in: *Computer Communications*

Received date: 28 July 2015
Revised date: 4 April 2016
Accepted date: 21 May 2016

Please cite this article as: Lizhi Peng, Haibo Zhang, Yuehui Chen, Bo Yang, Imbalanced Traffic Identification Using an Imbalanced Data Gravitation-based Classification Model, *Computer Communications* (2016), doi: 10.1016/j.comcom.2016.05.010

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Imbalanced Traffic Identification Using an Imbalanced Data Gravitation-based Classification Model

Lizhi Peng[a], Haibo Zhang[b], Yuehui Chen[a], Bo Yang*[a]

[a] *Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, 250022, P.R. China*
* *Corresponding author*
[b] *Department of Compute Science, University of Otago, Dunedin, New Zealand*

Abstract

As a basic method for monitoring the activities of Internet applications, traffic identification is very important for Internet management and security. Internet traffic data naturally exhibits imbalanced distributions, but this problem is rarely considered by the research community. Data gravitation-based classification (DGC) is a new classification model for handling imbalance data sets and we proposed an imbalanced DGC (IDGC) model in our previous study. In the present study, we developed an IDGC-based model to resolve imbalanced Internet traffic identification problems. First, we constructed six imbalanced traffic data sets from three original traffic data sets and we then extracted their early stage features according to the packet sizes. In identification experiments, we compared the performance of six standard algorithms, including DGC and four imbalanced algorithms with IDGC. The experimental results demonstrated that the standard classification models could achieve high accuracy with imbalanced traffic data, but their imbalanced performance was not as good, and their generalizability was also a problem. By contrast, IDGC performed very well and in a stable manner in terms of imbalanced performance measures, thereby demonstrating its effectiveness for imbalanced traffic identification.

Keywords: Data Gravitation, Imbalanced Data, Machine Learning, Traffic Identification

Email addresses: plz@ujn.edu.cn (Lizhi Peng), haibo@cs.otago.ac.nz (Haibo Zhang), yhchen@ujn.edu.cn (Yuehui Chen), yangbo@ujn.edu.cn (Bo Yang*)

Download English Version:

https://daneshyari.com/en/article/4954431

Download Persian Version:

https://daneshyari.com/article/4954431

Daneshyari.com