学校编码：10384

学号：X2010230658

厦门大学

硕士学位论文

基于细粒度特征的网络流量管理系统设计与实现

Design and Implementation of the Network Traffic Management System Based on Fine-grained Features

王存浩

指导教师：杨律青

专业名称：工程硕士(软件工程)

答辩日期：2012年11月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外，该学位论文为(　　　　　　　　　　　　)课题(组)的研究成果，获得(　　　　　　　　　　　　)课题(组)经费或实验室的资助，在(　　　　　　)实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人(签名):

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

(　　)1.经厦门大学保密委员会审查核定的保密学位论文，于　年　月　日解密，解密后适用上述授权。

(　　)2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年　　月　　日

摘　要

随着网络应用的蓬勃发展，特别是广泛应用的P2P技术，给网络的有效管理带来了很大的困难。为更好的实现网络的有效控制和管理，必须对对网络流量中各种应用进行准确的识别与分类。近几年来国内外学者围绕流量识别与分类进行了众多研究，并取得了相当可观的研究成果，但这些研究多是从宏观角度对网络流量进行建模分析，没有过多关注网络流内部随时间变化的动态特性和随用户行为变化的交互特性，且目前的网络流识别分类研究多为离线的方法，对网络流量的控制和管理作用有限。

针对这些不足，本论文采用了双向的动态网络流模型，充分考虑网络流内部的动态交互特性，细粒度的网络流量特征刻画弥补了目前网络流量微观特性研究方面的不足，为网络流量分析提供一种新的数据源与视角。

本文在双向动态网络流模型的基础上，采用细粒度的Packet-Level序列特征属性进行分析，研究不同流量特征对于网络流量分类与识别的重要性，研究网络流序列特征和网络应用之间的关联关系，在此基础上主要进行流量的实时识别与分类的研究。通过对P2P流量数据进行分析，得出关键包序列特征在区分P2P应用的重要性，因此将序列特征与动态规划结合起来，通过求取网络流序列特征与关键包序列特征的最长公共子序列的长度来实现P2P应用类型的实时标定。基于上述算法，本文设计并实现了基于细粒度特征的流量管理系统，主要实现在P2P网络流建立初期快速高效识别其应用类型的效果。

关键词：细粒度；动态网络流；序列特征

Abstract

Rapid development of network applications, particularly P2P technology brings difficulties to network management. Network flow identification and classification is very important for network security and management, which attracts many researchers all over the world in recent years, the researchers in this field have yielded fruitful achievements, but most of them focused on describing the trend of the aggregate volume from a macro perspective, with seldom attentions to the dynamic interactive features, and most of these achievements are offline classification, which play limited role in network management.

In order to compensate for deficiencies, the authors adopted the concept of dynamic network flow model, which describes the dynamic interactive features of network flow from a micro perspective, and also provides a new data source and perspective for network traffic analysis.

Based on two-way dynamic network flow model, the paper researches on the relationship between the sequence characteristics of network flow and network applications by analysing packet-level attributes from a fine-grained perspective. This research focuses on two aspects: the first is the classification of common application protocols, the second is real-time identification and classification of P2P traffic. Through the analysis of sequence characteristics, and combined it with visual computing, the binary frequency distribution constructed by packet-level sequence characteristics showed great differences, the paper achieved the classification of common application protocols by calculating a difference score. Key packets sequence characteristics is very important for the identification of P2P applications, the paper employed dynamic programming to identify the key packets from network flows. Experiments show that the classification algorithms are correct and effective. Based on the above algorithms, the paper designed and implemented traffic safety management system based on fine-grained features, mainly to achieve the real-time identification of P2P applications.

Keywords: Fine-grained Features; Dynamic Network Flow; Sequence Characteristics

参考资料

[1]中国互联网络发展状况统计报告[R/OL]. [2012-07-19]. http://www.cnnic.net.cn/research/bgxz/tjbg/201207/P020120719489935146937.pdf.

[2]Karagianis T, Papagiannaki D, Faloutsos M. BLINC: Multilevel Traffic Classification in the Dark [C]. Proc of ACM SIGCOMM. New York: ACM Press, 2005, 21: 229-240.

[3]Roughan M, Sen S, Spatscheck O, et al. Class-of-service Mapping for QoS: A Statistical Signature-based Approach to IP Traffic Classification[C]. Proc of ACM SIGCOMM Internet Measurement Conference. Taormina, Sicily, Italy. 2004: 135 ~148.

[4]Zuev D, Moore AW. Traffic Classification Using a Statistical Approach[C]. Proc of the 6th Passive and Active Measurement Workshop. Berlin:Springer, 2005: 321~324.

[5]Moore AW, Zuev D. Internet Traffic Classification Using Bayesian Analysis Techniques[C]. ACM SIGMETRICS. New York: ACM Press, 2005: 50-60

[6]Eerman J, Mahanti A, Arlitt M. Internet Traffic Identification Using Machine Learning Techniques[C]. Proc of 49th IEEE GLOBECOM, San Francisco, USA, 2006.

[7]Erman J, Arlitt M, Mahanti A. Traffic Classification Using Clustering Algorithms[C]. Proc of ACM SIGCOMM Workshop on Mining Network Data. Pisa, Italy. 2006.

[8]Wang R, Liu Y, Yang YX, et al. Solving the App-Level Classification Problem of P2P Traffic via Optimized Support Vector Machines[C]. Proc. of the 6th International Conference on Intelligent Systems Design and Applications. Jinan, Shandong, China. 2006: 534~539.

[9]徐鹏,林森．基于C4.5决策树的流量分类方法［J］．软件学报,2009-10-15.

[10]Nguyen T, Armitage G. Training on Multiple Sub-flows to Optimise the Use of Machine Learning Classifiers in Real-world IP Networks. Proc of the 31st IEEE LCN, Tampa, Florida. 2006: 14-17.

[11]Bernaille L, Teixeira R, Salamatian K. Early application identification[C]. Proc of Conference on Future Networking Technologies. Lisboa, Portugal. 2006.

[12]Haffner P, Sen S, Spatscheck O, et al. ACAS: Automated Construction of Application Signatures[C]. Proc of the 2005 ACM SIGCOMM Workshop on Mining Network Data, 2005: 197- 202.

[13]刘颖秋,李巍,李云春．网络流量分类与应用识别的研究［J］．计算机应用研究,2008,25(05)：210.

[14]邓河,阳爱民,刘永定．一种基于SVM的P2P网络流量分类方法［J］．计算机工程与应用,2008,44(14)：122-126.

[15]李江涛,姜永玲．P2P流量识别与管理技术［J］．电信科学,2005,21(3)：57-60.

[16]周世杰,秦志光,吴春江．对等网络流量检测技术研究［J］．电子科技大学学报,2007,36(6)：1333-1337.

[17]王锐,王逸欣,樊爱华,等．一种跨层P2P流量检测方法．计算机应用,2006,26(12)：30-32.

[18]边江．动态网络流分类方法研究［D］．西安：西安交通大学,2007.

[19]聂秦岭．基于动态网络流集合的网络用户行为分析［D］．西安：西安交通大学,2007.

[20]王盈．基于动态网络流的网络用户行为分析［D］．西安：西安交通大学,2008.