一种协议分类支撑系统

何丽[1]，龚俭[1]

（1.东南大学计算机科学与工程学院，南京，211189）

摘　要：标识出互联网流量所使用的应用层协议对于网络监测、网络管理等都具有重要的意义。CERNET 华东(北)地区网络中心网管系统 NBOS 的监测结果表明，一些非系统端口的使用有时会出现明显的增长，分析这些流量的应用层协议组成对优化网络管理很重要，为此需要一个支持协议分类和协议分析的基础平台。目前开源的协议分类系统如 openDPI 等只支持基于报文负载特征的分类方法，该方法有其固有的局限性，为了达到更好的分类效果，需要一个支持多样化分类方法的平台。为此本文提出构建一个协议分类支撑系统，文章首先介绍了协议分类的相关背景，随后介绍了协议分类支撑系统的总体设计，最后就系统的每个功能模块设计作详细介绍。

关键词：协议分类；流量识别；特征匹配；行为特征

A protocol classification support system

He Li[1], Gong Jian[1]

(1. School of Computer Science & Engineering, Southeast University, Nanjing, 211189)

Abstract: Application identification of network traffic is significant to the network monitoring and network management, etc. The network management system NBOS indicates that sometimes there is significant growth in the use of non-system port. Therefor a protocol classification and analysis system is needed for analysis the composition of the traffic for optimized the traffic managemant. Currently protocol classification system such as OpenDPI can only support classification method based on packet payload content. To improve the classification result a platform supporting diverse classification methods is needed. For that the paper proposed to construct a protocol classification support system. Firstly, the article introduces the background of protocol classification. Then the overall design of the system is present. Finally the paper describes each functional module of the system for details.

Key words: Protocol classification; Traffic identification; Feature matching; Behavioral characteristics

识别出当前网络流量中每条流所承载的应用层协议，可以实现对网络各类业务进行细粒度的监测与管理以及按需配置优化网络，对网络管理与网络监测非常重要。NBOS(Network Behavior Observation System)是 CERNET 华东(北)地区网络中心用于监控和管理 CERNET 网络服务质量和网络安全状态的新型精细化网管系统，该系统旨在设计和实现对网络流量的分析、异常发现和应急控制，限制网络异常流量对正常网络服务的影响，为用户提供网络服务质量监测和评估功能。NBOS 的监测结果表明一些非系统端口的使用会在某段时间内明显高于平均水平，这些非系统端口的流量占据总流量的比例上升，这些流量影响了正常的网络服务，为了更好的管理网络，需要诊断这些流量的形成原因进而优化网络配置，因此分析这些流量的应用层协议成份非常重要，为此需要一个支持协议分类和协议分析的平台。

目前开源的协议分类系统如 openDPI 等只支持基于报文负载特征的分类方法，该方法有其固有的局限性，为了达到更好的分类效果，需要一个支持多样化分类方法的平台。应用层协议分类方法主要经历了基于传输层端口、基于报文负载特征和基于流量行为特征几个阶段[1-2]。应用层协议分类初期采用基于传输层端口号的识别方法[3]，由于网络的高速化和协议的复杂化，基于传输层端口号的识别方法准确性降低，随之衍生出了各种新的识别方法。基于报文负载特征的识别方法是继端口号之后的研究方向，基于负载内容识别[4-6]首先要建立特征库，

作者简介：何丽，（1988-），女，硕士研究生，E-mail: lhe@njnet.edu.cn；龚俭，（1957-），男，教授，博导，E-mail: jgong@njnet.edu.cn.

在识别过程中采用特征匹配的方法识别应用协议，此方法在加密和隐私方面具有局限性，此后基于流量行为特征的识别方法[7-10]成为研究热点。基于流量行为特征的方法利用协议规范的不同所造成的流测度的差异区别各个协议，首先归纳出协议交互过程中在流/主机上表现出的不同行为特征再以此分类后续未知的流量，目前的研究热点是利用机器学习中的聚类或分类方法来标识流量。

上述应用层协议分类方法可划分为基于报文负载特征和基于流量行为特征两类，因此本文提出构建一个协议分类支撑系统，该系统以分析网络流量的应用层协议为目标，支持基于报文负载特征和基于流量行为特征两类分类方法，并且具有可扩展性，可在原来的基础上添加新的识别方法，此外系统提供特征分析功能，为未知类别流量的协议分析提供基础平台，最后为系统设计用户界面，实现通过界面使用和管理系统。

1 协议分类的目的

NBOS 是一个基于被动测量的新型网络管理系统，系统通过周期性地从路由器接收流记录数据，通过合理的网络区域划分，以适当的时间粒度为单位，准实时的分析被管单位不同网络互联级别的流量行为，评估被管单位和全网的服务质量，并将结果以 web 界面呈现。本系统是一个离线分析系统，用于 NBOS 系统监测总体流量行为时发现流量使用出现异常时，分析流量的应用层协议成份，为 NBOS 诊断异常流量的形成原因进而采取流量限制措施提供基础，本系统与 NBOS 的关系如图 1 所示。

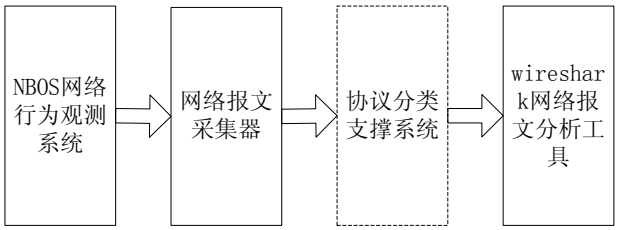

图 1 系统应用环境

当管理员发现 NBOS 监测到网络流量出现异常时，首先通过报文采集器在被管网络链路上捕获一定数量的原始流量，将捕获到的流量提供给协议分类支撑系统分析，协议分类支撑系统首先将可识别的应用层协议流量标识出，并提供自动分析的功能，基于已有的协议分类规则的测度特征计算未分类流量在这些测度上的取值，以便发现这些流量的测度规律，同时将未分类的流量以 pcap 格式存储，进而可以通过 wireshark 网络报文分析工具进行流量浏览与人工分析。

2 系统功能设计

本文提出的协议分类支撑系统旨在提供分析网络流量所使用的应用层协议的功能，该系统不仅支持基于负载内容的分类方法，还提供基于流量行为特征的分类方法，并且系统具有可扩展性，此外系统还提供对未知类型流量的分析功能，系统的功能结构图如图 2 所示。

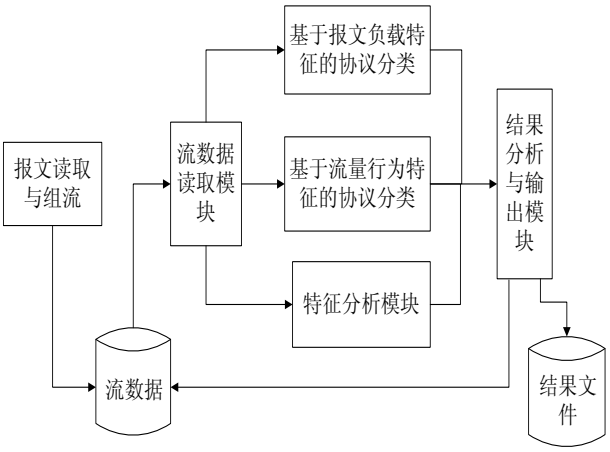

图 2 系统功能结构图

系统以全报文数据为输入，主要分为报文组流，协议分类和特征分析几个模块，系统的各个功能模块简述如下：

(1) 报文读取与组流：该模块读取输入的全报文数据，对报文进行组流处理，将报文以流为单位重新排列，生成流数据文件。流数据文件以流为单位存储，每个流的内容包括流记录信息与属于该流的所有报文。

(2) 流数据读取：该模块读取流数据文件，将读取的数据提供给协议分类模块或特征分析模块处理。

(3) 基于报文负载特征的协议分类：该模块按照已知的协议特征对读取的报文进行匹配，判断报文是否属于某种已知的协议，最后根据每条流所有报文的匹配结果决定该流的协议类别。该模块基于 OpenDPI 系统的协议分类功能设计，为每种可识别的协议添加插件，插件库具有可扩展性，可以更新已有的协议特征或添加新的

协议特征。

(4) 基于流量行为特征的协议分类：该模块根据协议的流量行为特征基于每条流的流记录信息判断该流的协议类别。该模块同样基于插件来实现，具有可扩展性。

(5) 特征分析模块：该模块基于已有的协议分类特征来分析未知类别流量在这些特征字段的取值，以期发掘出未知类别流量的特征。

(6) 结果分析与输出模块：该模块接收来自协议分类模块与特征分析模块三个模块的输出，将协议分类结果或特征分析结果保存到结果文件。

(7) 流数据：流数据文件为组流模块的输出，用于保存报文组流后的流信息，该文件是分类与分析模块的输入。

(8) 结果文件：结果文件为结果分析与输出模块的输出，用于保存处理模块得到的分析结果。

为实现系统工具化的需求，采用可视化的形式管理系统，系统基于 B/S 的结构，通过浏览器界面调用系统功能及管理系统配置。系统分为浏览器用户界面和后台功能模块两部分，系统的逻辑结构图如图 3 所示。

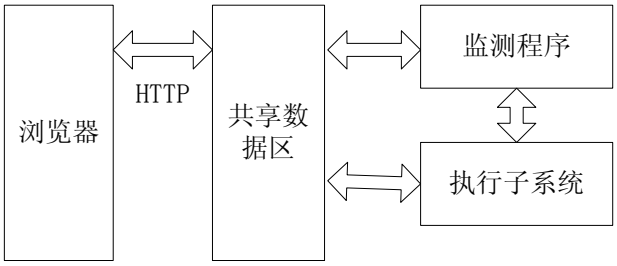

图 3 系统逻辑结构图

浏览器界面实现调用系统功能和管理系统配置的接口，共享数据区主要包括数据库和分析结果文件，发生系统功能调用时，浏览器向共享数据区的数据库插入新的任务，后台的监测程序通过监视数据库发现是否有新任务，监测到新任务便调用执行子系统去执行，执行子系统执行完成后将处理结果存入共享数据区，浏览器读取结果文件将分析结果反馈到用户界面。

3 系统总体结构

系统采用 B/S 架构，以 Web 方式呈现给用户，协议分类支撑系统的软件结构图如图 4 所示，系统包含用户界面、监测程序、协议分类子系统、特征分析子系统四个主要部分，图中列出了系统的功能模块及各个功能模块之间的调用关系。

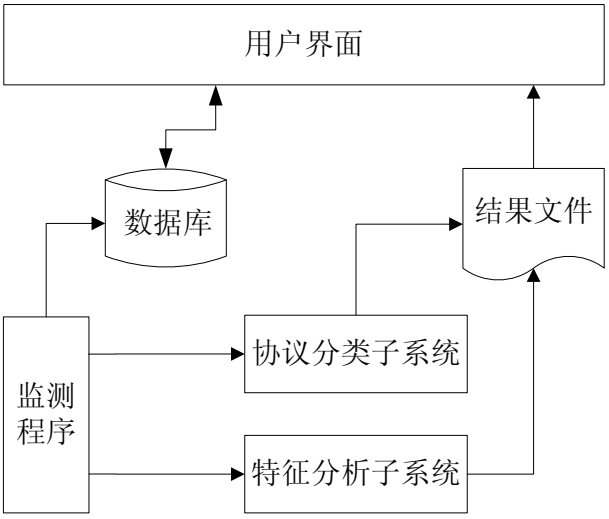

图 4 系统软件结构图

图中各个功能模块的设计将在下面几节分别介绍，这里对数据库和结果文件作简要说明。数据库存放测试数据，系统分类支持的协议以及协议插件信息，此外最重要的功能是通过任务表衔接用户界面与后台程序，达到通过用户界面调用系统功能模块的作用。结果文件是协议分类子系统和特征分析子系统的处理结果，以文本文件的形式存储，浏览器读取结果文件将处理结果反馈到用户界面。

3.1 用户界面

系统采用基于 B/S 结构的设计，通过浏览器进行可视化的系统管理，用户界面包含主页面，协议分类页面，特征分析页面，和配置管理页面四个部分，下面分别介绍。

系统主页面列出当前所有测试数据的信息，提供查看测试数据分类结果的功能，若该测试数据未分类则给出错误提示，否则将列出该测试数据的分类结果。

协议分类页面和特征分析页面分别用于调用协议分类子系统和特征分析子系统处理选中的测试数据，通过向数据库中插入新的任务实现功能调用，程序执行完成可通过浏览器查看得到的结果文件。

配置管理页面包含协议信息管理和测试数据管理两项内容，协议信息管理通过修改数据库中的基本表实现协议和协议插件的添加，测试数据管理实现测试数据信息的添加、删除和修改。

3.2 监测程序

监测程序采用轮询的方式工作，通过定期查询数据库任务表监视是否有新任务产生，当有新任务到达时，根据新任务的类型调用协议分类子系统或特征分析子系统执行当前任务，执行完成后将任务表中该任务的状态更新为完成状态，此后用户界面可以查看该任务的执行结果。

3.3 协议分类子系统

协议分类子系统包括组流模块和分类模块两个部分，组流模块读入报文生成流数据文件，分类模块执行协议分类过程生成带协议标识的流数据文件，下面对两个模块分别进行介绍。

3.3.1 组流模块

组流模块读取报文数据，按五元组流规范和超时约束对报文进行组流，组流过程将报文重新排列，生成流数据文件。为了操作的通用性，流数据文件与报文文件一样以 pcap 文件格式的形式存储，流数据文件的存储格式如图 5 所示。流数据文件以流为单位存储，每个流的内容包括该流的流记录与属于该流的所有报文。文件中的流按组流过程中每条流的结束时间的先后顺序排列，对于每条流，首先将该流的流记录构造成一个 pcap 文件的报文结构作为该流的第一个报文存储，为区别于其它报文该报文结构的时戳设置为 0，再将属于该流的所有报文存储在流记录之后，报文按到达顺序排列，所有流依次存储构成流文件。

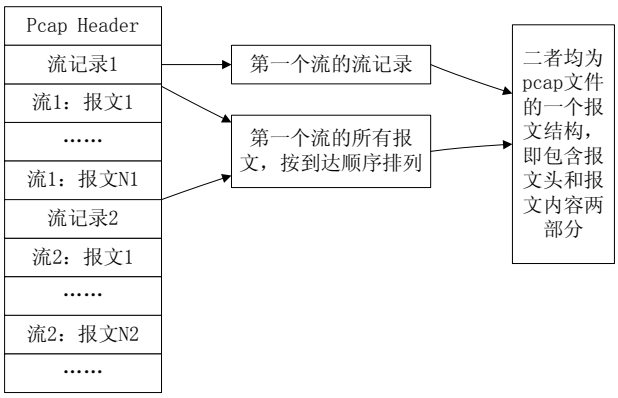

图 5 流文件存储格式

3.3.2 协议分类模块

协议分类模块以流数据文件为输入，通过匹配报文负载特征和流量行为特征，得到每条流的应用层协议类型，将每条流的协议类型写回流记录的协议类型字段生成带协议标识的流文件，并为未识别的流量单独生成一个流数据文件，最后将所有流的分类结果保存为结果文件。协议分类模块的处理流程如图 6 所示。

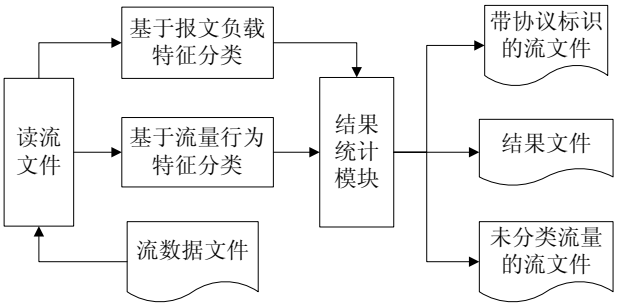

图 6 协议分类模块处理流程

首先读取流数据文件，若遇到数据时戳为 0 即该报文结构存储的流记录，则根据流记录中的报文数读取属于该流的所有报文，对每个报文基于负载特征分类，所有报文处理完毕再基于流量行为特征对该流进行分类，最后根据两种方法的分类结果标识流记录中的协议字段，将流记录构成报文写回文件。所有流处理完毕根据带协议标识的流数据文件统计分类结果，并重新排列流文件结构将其按分类之前的格式存储，同时将未识别的流量单独保存成一个流数据文件。

3.4 特征分析子系统

特征分析子系统以协议分类子系统生成的未知类别的流数据文件为输入，基于已有的协议分类规则的测度特征，计算未分类流量在这些测度上的取值，以期发现这些流量的测度规律。特征分析以流为单位，将每条流的分析结果生成结果文件，特征分析子系统的处理流程如图 7 所示。

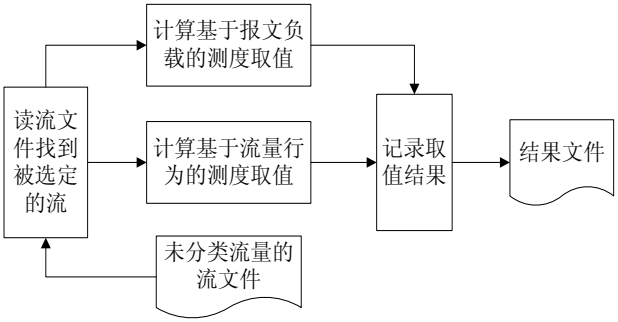

图 7 特征分析模块处理流程

特征分析以流为单位，首先读取未分类流量的流数据文件获得选中的流，针对基于报文负载的协议分类规则计算该流的每个报文在规则对应测度上的取值，针对基于流量行为的协议分类规则计算机该流在规则对应测试上的取值，记录取值结果生成分析结果文件。

4 小结

本文分析了网络管理中应用层协议分类的需求，提出了一个协议分类支撑系统的设计方案，对系统的总体设计和功能模块设计作了详细的介绍，该系统支持协议分类和特征分析的功能，具有良好的可扩展性。

参考文献

[1] 陈亮, 龚俭, 徐选. 应用层协议识别算法综述[J]. 计算机科学, 2007,34(7):72-75.

[2] 余浩, 徐明伟. P2P 流检测技术研究综述[J]. 清华大学学报(自然科学版), 2009,49(4):616-620.

[3] IANA.PORT NUMBERS[S]. http://www.iana.org/assignments /port-numbers. 2011-12-22.

[4] Ipoque GmbH. The open source deep packet inspection engine[CP/OL]. http://www.opendpi.org/.

[5] quadong, sommere. SourceForge.net: Linux layer 7 packet classifier[CP/OL]. http://sourceforge.net/projects/l7-filter/. 2009-5-28.

[6] Subhabrata Sen, Oliver Spatscheck, Dongmei Wang. Accurate, scalable in-network identification of p2p traffic using application signatures[C]. In: Proc.of the 13th international conference on World Wide Web. New York, USA, 2004:512-521.

[7] Thomas Karagiannis, Andre Broido, Michalis Faloutsos, Kc claffy. Transport layer identification of P2P traffic[C]. In: Proc.of IMC'04, Taormina, Sicily, Italy, 2004:25-37.

[8] Jeffrey Erman, Martin Arlitt, Anirban Mananti. Traffic Classification Using Clustering Algorithms[C]. In: Proc. of ACM SIGCOMM'06. MineNet Workshop. Pisa, Italy, 2006: 281-286.

[9] Zander S, Nguyen T, Armitage G J. Self Learning IP Traffic Classification Based on Statistical Flow Charact eristics [C] . In: PAM2005. Boston, USA, 2005: 325-328.

[10] Zuev D, Moore A W. Traf fic Classification Using a Statistical Approach[C] . In: PAM2005. Boston, USA, 2005: 321- 324.