

文章编号: 2095-4980(2020)05-0918-06

一种基于 Hadoop 平台 CloudSVM 的网络流量分类方法

邓 河, 唐一韬, 贺宗梅, 袁爱平

(长沙民政职业技术学院 软件学院, 湖南 长沙 410000)

摘 要: 大规模的 netflow 训练数据集是构建高质量、高稳定网络流量分类器的必然要求。但随着网络流特征维数的提高和数据集规模的扩大, 无论是网络流的分析处理还是基于支持向量机(SVM)的分类器模型的训练, 都无法在有效的时间内得到有效的处理结果。本文基于 Hadoop 云计算平台, 采用 MapReduce 技术对 SVM 网络流量分类器进行分布式学习和训练, 构建 CloudSVM 网络流量分类器。通过对来自校园网出口镜像的近 2 T 的大规模网络流量的跟踪文件的分布式存储和处理, 对抽取的样本数据集进行分类, 实验验证了基于 Hadoop 平台分布式存储和并行处理大规模网络数据集的高效率性, 也验证了 CloudSVM 分类器在不降低分类准确度的情况下可以快速收敛到最佳, 并随着大规模网络流样本的增加, SVM 分类器训练的时间趋近平稳。

关键词: 网络流量分类; Hadoop 平台; CloudSVM 分类器

中图分类号: TP393

文献标志码: A

doi: 10.11805/TKYDA2019406

A network traffic classification method based on Hadoop platform CloudSVM

DENG He, TANG Yitao, HE Zongmei, YUAN Aiping

(School of Software, Changsha Social Work College, Changsha Hunan 410000, China)

Abstract: Large-scale net flow training data sets are inevitable requirements for building high-quality, highly stable network traffic classifiers. However, with the increase of the feature dimension of the network stream and the expansion of the data set size, neither the analysis processing of the network stream nor the training of the classifier model based on Support Vector Machin(SVM) can obtain effective processing results in effective time. A distributed and parallel large-scale network flow based on Hadoop cloud computing platform is proposed. Distributed learning and training of SVM network traffic classifier is implemented by MapReduce technology on Hadoop cloud computing platform, and CloudSVM network traffic classifier is constructed. Through the distributed storage and processing of trace files of large-scale network traffic from the campus network export mirror, the sample data sets are classified, and the distributed storage and parallel processing of large-scale network data based on Hadoop platform is experimentally verified. The high efficiency of the set also verifies that the CloudSVM classifier can quickly converge to the best without reducing the accuracy of the classification, and with the increase of large-scale network flow samples, the training time of the SVM classifier is approaching constant.

Keywords: network traffic classification; Hadoop; CloudSVM

基于机器学习方法的网络流分类, 是对基于统计特征所表征的网络流(netflow)数据样本, 采用机器学习方法训练并构建分类器, 对网络流量进行分类的一种新的分类方式。这种分类方式日趋深入到应用中, 如 Arivudainambi D 等使用主成分分析和人工神经网络对恶意流量进行检测和分类^[1]; 邹腾宽也用基于机器学习的方式对网络背景流量进行分类^[2]。这种基于机器学习的网络流量分类的准确率很大程度上依赖训练样本集的大小以及特征的维度。较大的训练样本集和较高维度的特征集可以提高分类器的分类准确度。传统的网络流处理和分类器的训练都是基于单机进行, 如文献[3]采用深度学习对网络流量进行分类; 莫遥等^[4]采用

收稿日期: 2019-10-20; 修回日期: 2019-11-29

基金项目: 湖南省教育厅科研资助项目(15C0081); 湖南省教育厅科研资助项目(14C0064); 湖南省教育厅科研资助项目(19C0103)

作者简介: 邓 河(1977-), 男, 硕士, 讲师, 主要研究方向为机器学习、网络流量分类、情感文本分类、数据挖掘。email:57898971@qq.com

决策树算法,采集 P2P 和非 P2P 流量的 256 维统计特征,在多种流量混杂的情况下识别出 P2P 流量;文献[5]中采用单机离线处理交换机出口镜像的跟踪文件。但随着互联网技术的快速发展和不断更新的实时应用要求,海量的网络流量数据的存储和分析处理给传统的计算处理方式带来重大的挑战,基于单机方式的网络流处理和分类已达不到快速、实时和可伸缩性的需求。Hadoop 平台的出现,解决了存储、处理和分析海量数据(大规模级, TB 级以上)的效率问题。Hadoop 平台本质是一种云计算的实现,重点包括分布式文件系统(Hadoop Distributed File System, HDFS)和 MapReduce 编程模型^[6]。通过 Hadoop 的分布式并行计算,将大大提高网络流量特征的计算以及网络流量分类器的训练速度。如, Y Lee 等提出了一个可扩展的基于 Hadoop 的并行数据包处理器,分析大型数据包跟踪文件^[7]; Y Lee 等设计了一个基于 Hadoop 的流量监控系统,该系统以可扩展的方式对 IP 量级的数 TB 的 Internet 流量执行 IP, TCP, HTTP 和 netflow 分析。使用 200 个结点的测试平台进行实验,通过 IP 和 HTTP 层分析 MapReduce 作业,为 5 TB 文件实现了 14 Gbps 的吞吐量,实现了可伸缩的 Internet 流量测量和分析^[8]。Chhabra 等基于 Hadoop 分布式文件系统环境 MapReduce 框架中实施监督机器学习(基于随机森林的决策树)算法,证实了所提出的框架在网络取证中的优越性,其平均准确度为 99.34%^[9]。Ying Zou 设计了一个基于 Hadoop 的移动互联网流量分析系统,对移动互联网流量数据的 URL 字段进行分类并在 URL 站点上挖掘,大大减少了分析处理时间,并扩展了用户在流量信息中的网站偏好维度^[10]。Weixiang Xu 等设计了不确定交通数据流并行连续查询算法,应用滑动窗口模式在数据流环境中实现数据的接收和缓存,以适应 Hadoop 分布式结构的 MapReduce 计算框架,实现了分布式环境中不确定交通数据流的并行连续查询^[11]。董欣基于 Hadoop 技术,针对网络流量大数据,设计了一个以分布式 HDFS 做为数据的存储媒介,以 Hbase 提供交互式查询的分析系统^[12]。曹佳豪建立了一个基于 Hadoop 的报文数据挖掘系统,将采集到的海量报文数据清洗并格式化后存储到 HDFS,然后根据 Apriori 算法进行关联规则挖掘^[13]。王勇等采用序列前向搜索策略,通过多层 MapReduce,实现不断选取分类能力最好的特征^[14]。梁文国等使用 Spark 中的 SVM 算法,对二分 SVM 子分类器进行训练,最后组合成并行多分类器^[15]。满蔚仕等结合网格法与粒子群算法,基于 Hadoop 平台实现了一种分布式 SVM,提高了分类准确度^[16]。吴云蔚等利用 Hadoop 平台对 SVM 进行分布式参数寻优,解决对实时性要求较高的系统在寻优过程中耗时过长的的问题^[17]。大部分的研究者对基于 Hadoop 平台下的分类器的训练和分类过程中的参数调整以及分类准确度做了一些研究,但缺少对大规模原始报文数据的分布式处理及在分类过程中的流特征向量的动态筛选过程。本文旨在对大规模原始报文数据采用分布式存储和处理,利用 CloudSVM 在云计算环境中找出与最优分类函数有关的支持向量,快速建立网络流分类器,实现大规模网络流的快速分类。

1 网络流的分布式存储和特征计算

1.1 网络流的分布式存储

报文(Packets)是网络通信的基本单位,也是基于机器学习方法网络流量分析的最小单位。报文根据相同的五元组(即双向通信的 IP 地址、端口号和 IP 协议)和其他一些约定信息形成流(flow),这些约定信息如包括流的 Idle Timeout<60 s 等;再对网络流的无关协议和端口的一些特征进行统计,从而形成特征网络流的特征向量;最后根据应用类型对流进行分类。本文采用文献[5]中建立的一个流应用类型标识系统对流应用进行标识,这个标识系统需基于有效载荷对流进行分析和标识^[18]。本文的报文数据通过截取报文的前 128 个字节的内容,以 Libpcap(*.dmp)格式存储,捕获报文格式如图 1 所示。过程中采集到的数据通过 Hadoop 集群环境分发到 HDFS 的块上去。其中 dmp 文件头需在分发到 HDFS 前进行预处理,实验中将 dmp 文件头信息以(Key, Value)的方式存储至 Hadoop 的 NameNode 上。

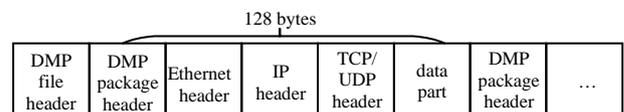


Fig.1 Schematic diagram of Libpcap (*.dmp) file format
图 1 Libpcap(*.dmp)文件格式示意图

HDFS 块默认大小为 64 MB,这样,每个 Block 存储 $64 \times 1024 \times 1024 / 128 = 524288$ 个完整的数据报文信息。

1.2 流特征选取

网络流量的特征选择决定最终流量分类器的分类质量,太多会影响分类器的构建速度,太少又可能降低分类的准确度。MOORE A W 等在文献[19]中提出了 248 个可用于流量分类的统计特征属性。本文对文献[17]中挑选的 37 个流的候选特征进行统计计算。基于这 37 个特征来构建分类器,特征如表 1 所示,这些特征具有较好的可分性。表中的特征值都从两个方向计算,除了持续时间特征。

表 1 分类中采用的网络流特征
Table 1 Network flow characteristics used in classification

feature	description
packets	total number of packages
bytes	total package size
PktLen	package minimum, maximum, average length and mean square error
inter-arrived time	time of mutual arrival, including minimum time, maximum time, average time and mean square error
active time	minimum, maximum, average activity time and mean square error
idle time	minimum, maximum, average idle time and mean square error
duration	duration

1.3 基于 MapReduce 的流分析和特征计算

MapReduce 是一个并行的计算框架，可用于大规模的网络流数据集的并行运算。流数据集的操作将分发给多个分结点来并行操作完成。分为两个工作过程：第一个是 Map 过程计算；第二个过程则进行 Reduce 工作。将捕获并处理的网络报文跟踪文件分布式部署在 HDFS 上。在 Map 阶段完成报文信息的提取，在 Reduce 阶段进行流的解析。具体实现过程如图 2 所示。在 Map 阶段，将分布式存储在 HDFS 上的 DMP 文件二进制数据块作为输入。每个 Map 作用在对应的 Block 上，依次读取 128 个字节的报文数据并解析出报文字段。Map 阶段的输出是<key,value>形式的文本行数据。Key 用流的五元组按顺序组成的字符串做为键，值即为报文的字段信息，包括包长、时间戳、标志位等。字段信息间用逗号分隔。多层 Reducer 结构的第一层把 Map 解析的报文字段信息根据网络流形成规则进行流的形成、流特征的计算统计，第二层开始的 Reduce 过程将上一层 Reduce 过程中形成的网络流按 Key 进行合流(根据时间戳)，并统计计算相关特征。

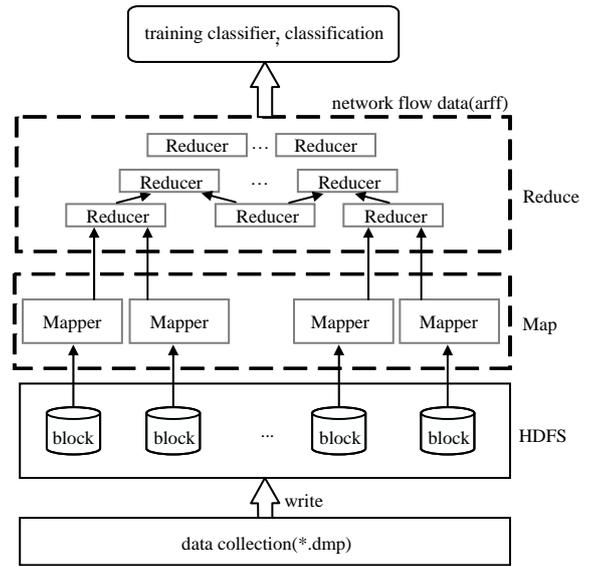


Fig.2 MapReduce algorithm processing of network flow
图 2 网络流 MapReduce 算法处理

2 CloudSVM 训练

基于 SVM 对网络流量进行分类识别可有效减少样本分布，降低冗余度，避免过度拟合现象发生^[20]。面向大数据的 SVM 算法，需要将原始空间映射到高维空间，再好的核函数形成的高维空间都将占用大量的内存空间。随着数据的增长，内存消耗呈平方上升。目前 SVM 采用 SMO 算法进行训练^[21]，它将最终求解分成 N 个小问题的求解，由 N 个小问题的求解得到最终解。但单机 SMO 算法对于大规模的数据集，收敛速度非常慢。

传统方法中，会在预先配置好的 Intranet/Internet 环境上训练分布式 SVM 算法，以找出最佳分类器。对于大型数据集，这些方法非常复杂且成本很高。因此，Catak F 提出了一种在云计算环境中使用 MapReduce 技术针对分布式机器学习应用程序的方法，称为 Cloud SVM 训练机制(CloudSVM)^[22]。CloudSVM 训练的过程就是在云计算环境中找出与最优分类函数有关的支持向量，即让 SVM 收敛到最佳分类器功能。网络流量是一个典型的大规模数据集，CloudSVM 算法对于大型训练数据很适用：当迭代为 3~5 时，所有数据集的测试精确度值将达到最高值；如果迭代大小增加，则测试精确度的值将进入稳定状态。因此，本文选用 CloudSVM 对网络流样本进行训练。将大规模的网络流训练样本数据集在 Hadoop 云计算平台上进行分块，由一个大的训练数据集划分为多个小的子训练数据集，把每个子训练数据集上传至 Hadoop 的子结点上，在结点上对每个子训练数据集进行单独训练，收集好每个结点上得到的支持向量集，最后，将所有子结点上训练后获得的支持向量合并，通过进一步的训练得到网络流量的 SVM 分类模型。训练过程如图 3 所示，网络流训练集被分成多个子集，每个子集

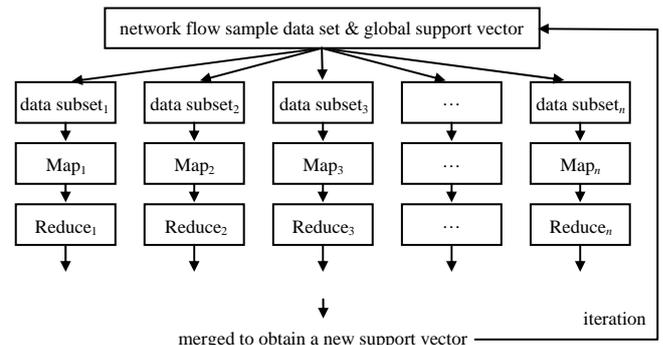


Fig.3 Diagram of CloudSVM training step
图 3 CloudSVM 训练步骤过程图

在 Map 步骤中进行单独训练并获得支持向量，然后再与全局支持向量相结合。在 Reduce 步骤中，评估训练数据的合并子集，得到新的支持向量与 Reduce 步骤中的全局支持向量相结合。

具体步骤如下：

步骤 1 设置 $t=0$, V' 为空集, t 为迭代次数, V 为原始支持向量集合。随机选择 200 个网络流样本集, 在 Hadoop 的主结点上训练, 通过训练得到基于 SVM 的网络流量初始分类器。

步骤 2 迭代开始, $t=t+1$ 。

步骤 3 在每个子结点计算机上读取全局的支持向量, 将其与数据子集合并, 并通过 SVM 分类器进行分类, 经过计算获得新的支持向量。

步骤 4 计算以后将所有支持向量合并, 并存储为全局支持向量。

步骤 5 如果 $h^t=h^{t-1}$ 满足迭代终止条件, 则停止; 否则, 跳到步骤 2 继续进行。

3 试验结果及分析

3.1 实验环境

实验的网络流量数据采集于长沙民政职业技术学院图书信息中心交换机。通过采集工具软件抓包 24×2 h, 截取捕获的数据将近 2 TB。基于 Hadoop 的网络流量处理及分类器训练分别在 3 种不同规模大小及不同硬件配置环境下实验, 实验环境如表 2 所示。

集群系统里设置 1 台为名称结点(NameNode), 即主结点, 负责网络流的分块及映射关系管理等。另外的结点为数据结点(DataNode), 即分结点, 负责网络流数据的具体存储及流处理工作。利用 1 台分结点负责报文跟踪文件的接收, DMP 文件处理、网络流合成、特征的计算以及 SVM 分类器的训练由整个集群系统完成。

3.2 网络流特征统计处理分析

在实验中, 将数据划分成五个量级分别在四种不同环境下进行网络流量数据的特征统计处理, 其处理的耗时如表 3 所示。

表 3 网络流单机处理与 Hadoop 集群处理耗时
Table3 Time-consuming of single network processing and Hadoop cluster processing

DMP size	time-consuming(stand-alone)/s	time-consuming/s		
		cluster 1	cluster 2	cluster 3
2 GB	103	216	485	289
100 GB	3 540	1 164	2 762	1 037
500 GB	10 254	1 513	3 899	1 223
1 024 TB	21 653	2 378	5 321	1 370
2 048 TB	47 936	3 254	6 763	1 554

从耗时表中可以得出以下结论：

1) 单机在统计网络流特征过程中的耗时随数据规模的扩大基本呈线性增长。当处理 2 GB 的较小数据时, 其处理时间反优于 Hadoop 集群方式, 这是源于 Hadoop 集群系统的系统初始化时间、中间文件的产生、网络之间的通讯和传输等问题的存在, 使得 Hadoop 集群并没有发挥其优势。

2) 3 个不同规模和不同硬件配置的 Hadoop 集群的处理时间都不相同, 但总体上随网络流量数据规模的扩大并没有呈线性增长, 时间变化较为平稳。尤其是当数据规模达到 100 GB 以上时, 多结点的集群(集群 3)的处理时间相比结点少的集群(集群 1)要优秀很多, 完全可以抵消掉多结点之间的网络通信的耗时。

3) 在相同的集群规模中, 硬件的配置对计算的耗时影响较大, 尤其是 CPU 硬件。

3.3 分类器训练分析

为分析不同规模数据集下的 CloudSVM 训练结果, 采用了 2 个不同的网络流数据集: 一个是 Moore 教授等采集的网络流量实验数据集, 每个流用 248 个特征进行表征^[19], 标为 Moore_set; 另一个是通过 Hadoop 集群系统对网络流分析和处理的网络流数据集, 标为 Self_set。数据集情况如表 4 所示。

表 4 实验数据集
Table4 Experimental data set

type	flow size(Moore_set)	flow size(Self_set)
WWW	78 410	80 206
FTP	5 260	3 871
MAIL	20 325	18 750
P2P	21 068	2 5205
summary	125 063	128 032

从表 4 中选择了常见的 4 类网络流数据做为 SVM 分类器训练和学习。2 个数据集分别约为 12 万条网络流样本。实验数据集中的训练样本占 70%，测试样本占 30%。在分类的精确度方面，采用精确率、召回率、F 测试值 3 个指标来评估分类器的质量。

实验中对比了单机 SVM(SingleSVM)和 CloudSVM 的分类精确度的情况，结果如表 5、表 6 所示。

表 5 Self_set 数据集分类精确度比较

test method	index		
	precision	recall	F-measure
SingleSVM	0.928	0.956	0.942
CloudSVM	0.922	0.949	0.935

表 6 Moore_set 数据集分类精确度比较

test method	index		
	precision	recall	F-measure
SingleSVM	0.941	0.953	0.947
CloudSVM	0.945	0.948	0.946

实验结果显示，CloudSVM 在 2 个数据集上的分类效果非常好，跟 SingleSVM 的分类效果相差不大。但它们在训练分类器所使用的时间却截然不同。训练耗时如图 4 所示。在 SVM 分类器的训练时间上，SingleSVM 的训练随网络流量数据集的扩大呈线性上升，Moore_set 数据集因为特征值较多，通过 SingleSVM 训练时，收敛很慢，耗时非常大；基于 CloudSVM 分类器的训练随网络流量数据集的扩大，训练时间比较平稳，Moore_set 数据集虽然特征值较大，但在迭代训练时，收敛的速度比较快，所以同特征值少的 Self_set 数据集的训练时间非常接近。通过实验发现，CloudSVM 在进行 SVM 分类器训练时，迭代 3~5 次就能迅速找到支持向量。在增大特征维度的情况下，同样可以快速收敛。

图 4 中的 CloudSVM 是在集群 1 环境中进行训练。为了解在不同规模的集群条件下 SVM 分类器的训练表现，分别在集群 1 和集群 3 环境下对 SVM 分类器进行训练，不同规模的数据集的耗时对比如图 5 所示。大规模的数据集随训练数据样本的增加，训练时间的增长更加平滑。

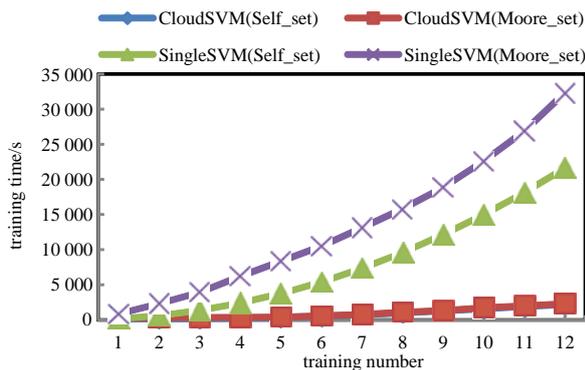


Fig.4 Comparison of training time between SingleSVM and CloudSVM
图 4 SingleSVM 和 CloudSVM 训练耗时对比

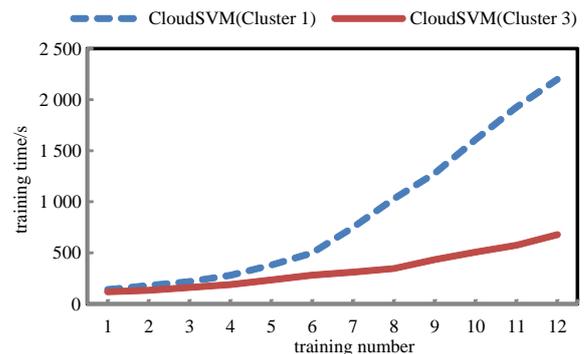


Fig.5 Comparison of training time in different clusters
图 5 不同集群下的训练耗时对比

4 结论

基于机器学习方法的大规模网络流量数据集的实时或准实时的正确分类，对传统的单机处理计算模式提出了严峻的挑战。随着云计算平台 Hadoop 的出现和不断成熟，基于 Hadoop 的分布式存储和采用 Map,Reduce 的云计算对大规模的网络流量数据集的处理和分类提供了可行的操作。本文提出的基于 Hadoop 的 CloudSVM 的网络流量分类就是借助 Hadoop 云计算平台，采用分布式并行计算的思想用于网络流的采集处理以及 SVM 分类器的训练，在不损失分类准确性的前提下大大提高了网络流从处理到分类器训练的时间。通过实验证明，基于 Hadoop 平台下的 CloudSVM 的网络流量处理和分类具有良好的可伸缩性和扩展性。当然，这些研究都是基于网络流在标识的条件下进行分类器训练和分类的。但现实中的网络流分类有时需要新的网络环境条件下进行无监督学习的分类，这将是接下来的研究方向。

参考文献：

- [1] ARIVUDAINAMBI D, VARUN KUMAR K A, SIBI C S, et al. Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance[J]. Computer Communications, 2019(147):50-57.
- [2] 邹腾宽, 汪钰颖, 吴承荣. 网络背景流量的分类与识别研究综述[J]. 计算机应用, 2019, 39(3):802-811. (ZOU Tengkuang, WANG Yuying, WU Chengrong. A review of classification and recognition of network background traffic[J]. Journal of Computer Applications, 2019, 39(3):802-811.)

- [3] SHU Junhua,JIANG Jiang,SUN Jingxuan. Network traffic classification based on deep learning[J]. Journal of Physics: Conference Series, 2018,1087(6):062021-1-5.
- [4] 莫遥,梁铸,吴波,等. 基于传输层特征和统计特征的 P2P 流量识别[J]. 太赫兹科学与电子信息学报, 2019,17(1):112-117. (MO Yao,LIANG Zhu,WU Bo,et al. P2P traffic recognition based on transmission layer characteristics and statistical characteristics[J]. Journal of Terahertz Science and Electronic Information Technology, 2019,17(1):112-117.)
- [5] 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(14):122-126. (DENG He,YANG Aimin,LIU Yongding. A SVM-based P2P network traffic classification method[J]. Computer Engineering and Applications, 2008(14):122-126.)
- [6] WHITE Tom. Hadoop:the definitive guide[M]. Sebastopol,California,USA:O'Reilly Media,Inc, 2009.
- [7] LEE Y,KANG W,LEE Y. A Hadoop-based packet trace processing tool[C]// Proceedings of the 3rd International Conference on Traffic Monitoring & Analysis. Berlin:Springer-Verlag, 2011:51-63.
- [8] LEE Y. Toward scalable internet traffic measurement and analysis with Hadoop[J]. Computer Communication Review:A Quarterly Publication of the Special Interest Group on Data Communication, 2013,43(1):5-13.
- [9] CHHABRA G S,SINGH V,SINGH M. Hadoop-based analytic framework for cyber forensics[J]. International Journal of Communication Systems, 2018,31(15):e3772.1-e3772.17.
- [10] ZOU Ying. Data analysis and processing of massive network traffic based on cloud computing and research on its key algorithms[J]. Wireless Personal Communications:An International Journal, 2018,102(4):3159-3170.
- [11] XU Weixiang,LI Jiaojiao. An improved algorithm for clustering uncertain traffic data streams based on Hadoop platform[J]. International Journal of Modern Physics B, 2019,33(19):1950203.
- [12] 董欣. 基于 Hadoop 和 HBase 的网络流量分析系统设计与实现[D]. 大连:大连理工大学, 2018. (DONG Xin. Design and implementation of network traffic analysis system based on Hadoop and HBase[D]. Dalian,China:Dalian University of Technology, 2018.)
- [13] 曹佳豪. 基于 Hadoop 的报文数据挖掘研究与应用[D]. 武汉:武汉邮电科学研究院, 2018. (CAO Jiahao. Research and application of message data mining based on Hadoop[D]. Wuhan,China:Wuhan Institute of Posts and Telecommunications, 2018.)
- [14] 王勇,龙也,陶晓玲,等. 基于多层 MapReduce 的混合网络流量分类特征选择方法[J]. 桂林电子科技大学学报, 2016,36(2):123-128. (WANG Yong,LONG Ye,TAO Xiaoling,et al. Feature selection method for traffic classification in hybrid network based on multi-layer MapReduce[J]. Journal of Guilin University of Electronic Technology, 2016,36(2):123-128.)
- [15] 梁文国,王勇,俸皓. 基于并行 DAGSVM 的网络流量分类方法[J]. 计算机工程与设计, 2018,39(2):316-319,335. (LIANG Wenguo,WANG Yong,FENG Hao. Classification of network traffic based on parallel DAGSVM[J]. Computer Engineering and Design, 2018,39(2):316-319,335.)
- [16] 满蔚仕,吉元元. Hadoop 平台分布式 SVM 算法分类研究[J]. 计算机系统应用, 2017,26(8):141-146. (MAN Weishi,JI Yuanyuan. Classification of distributed SVM algorithms on Hadoop platform[J]. Application of Computer Systems, 2017,26(8):141-146.)
- [17] 吴云蔚,宁芊. 基于 Hadoop 平台的分布式 SVM 参数寻优[J]. 计算机工程与科学, 2017,39(6):1042-1047. (WU Yunwei,NING Qian. Optimization of distributed SVM parameters based on Hadoop platform[J]. Computer Engineering and Science, 2017,39(6):1042-1047.)
- [18] 邓河. 基于机器学习方法的网络流量分类研究[D]. 长沙:湖南工业大学, 2009. (DENG He. Research on network traffic classification based on machine learning method[D]. Changsha,China:Hunan University of Technology, 2009.)
- [19] MOORE A W,ZUEV D. Internet traffic classification using Bayesian analysis techniques[J]. ACM SIGMETRICS Performance Evaluation Review, 2005,33(1): 50-60.
- [20] 邓绯. 基于支持向量机网络流量识别的研究[J]. 兰州文理学院学报(自然科学版), 2019,33(2):62-66,81. (DENG Fei. Research on network traffic identification based on support vector machine[J]. Journal of Lanzhou University of Arts and Science(Natural Science Edition), 2019,33(2):62-66,81.)
- [21] 王平,毛剑琴. 支持向量机训练算法及其应用[J]. 太赫兹科学与电子信息学报, 2005,3(4):309-314. (WANG Ping, MAO Jianqin. Support vector machine training algorithm and its application[J]. Journal of Terahertz Science and Electronic Information Technology, 2005,3(4):309-314.)
- [22] CATAK F O,BALABAN M E. CloudSVM:training an SVM classifier in cloud computing systems[C]// Joint International Conference on Pervasive Computing and the Networked World. Berlin:Springer-Verlag, 2013:57-68.