Journal of Computer Applications
计算机应用,2020,40(7):1970 - 1976
ISSN 1001-9081
CODEN JYIIDU
2020-07-10
http://www.joca.cn

文章编号:1001-9081(2020)07-1970-07
DOI:10.11772/j.issn.1001-9081.2019122063

基于大小突发块划分的微信支付行为识别模型

梁登高[1],周安民[1],郑荣锋[2*],刘　亮[1],丁建伟[3]
(1. 四川大学 网络空间安全学院,成都 610065; 2. 四川大学 电子信息学院,成都 610065;
3. 中国电子科技集团公司第三十研究所 保密通信重点实验室,成都 610041)
(*通信作者电子邮箱 qswhs@foxmail.com)

摘　要:针对微信红包与转账功能被用于红包赌博、非法交易等违法活动,且现有的研究工作难以识别微信中收发红包与转账行为的具体次数,以及存在低识别率和高资源消耗的问题,提出了一种划分大、小流量突发块的方法来提取流量特征,从而对收发红包与转账行为进行有效识别。首先,利用收发红包与转账行为流量的突发性,设定大突发时间阈值将这类行为的流量突发块分隔开;然后,针对收发红包与转账行为由多次连续的用户操作组成的特性,设定小突发阈值将流量块进一步细化为小突发块;最后,综合大突发块中各个小突发块的特征,得到最终的特征。实验结果显示,该方法在时间效率、空间占用率、识别准确率、算法普适性等方面普遍优于微信支付行为识别方面的现有研究,平均准确率最高可达97.58%。真实场景的测试结果表明,所提出的方法基本能准确识别出一段时间内用户收发红包与转账行为的次数。

关键词:微信;红包与转账;流量突发块;加密流量;机器学习
中图分类号:TP393.08　　**文献标志码**:A

WeChat payment behavior recognition model based on division of large and small burst blocks

LIANG Denggao[1], ZHOU Anmin[1], ZHENG Rongfeng[2*], LIU Liang[1], DING Jianwei[3]
(1. *College of Cybersecurity, Sichuan University, Chengdu Sichuan* 610065, *China*;
2. *College of Electronics and Information Engineering, Sichuan University, Chengdu Sichuan* 610065, *China*;
3. *Science and Technology on Communication Security Laboratory, The* 30*th Research Institute of China Electronics Technology Group Corporation, Chengdu Sichuan* 610041, *China*)

Abstract: For the facts that WeChat red packet and fund transfer functions are used for illegal activities such as red packet gambling and illegal transactions, and the existing research work in this field is difficult to identify the specific numbers of sending and receiving red packets and fund transfers in WeChat, and there are problems of low recognition rate and high resource consumption, a method for dividing large and small burst blocks of traffic was proposed to extract the characteristics of traffic, so as to effectively identify the sending and receiving of red packets and the transfer behaviors. Firstly, by taking advantage of the suddenness of sending and receiving red packets and fund transfers, a large burst time threshold was set to define the burst blocks of such behaviors. Then, according to the feature that the behaviors of sending and receiving red packets and fund transfers consist of several consecutive user operations, a small burst threshold was set to further divide the traffic block into small bursts. Finally, synthesizing the features of small burst blocks in the big burst block, the final features were obtained. The experimental results show that the proposed method is generally better than the existing research on WeChat payment behavior recognition in terms of time efficiency, space occupancy rate, recognition accuracy and algorithm universality, with an average accuracy rate up to 97.58%. The test results of the real environment show that the proposed method can basically accurately identify the numbers of sending and receiving red packets and fund transfers for a user in a period of time.

Key words: WeChat; red packet and fund transfer; traffic burst; encrypted traffic; machine learning

0　引言

微信2018年月活跃用户量高达10.8亿,且年龄覆盖率高,其中55岁以上的用户占6 300万[1]。微信中收发红包功能是互联网时代下中国传统红包的新表现形式,在2014年一经推出,便深受民众喜爱,使用人数急剧增长。2019年春节前五天,在线收发红包人次达到8.23亿[2]。微信中转账功能为日常生活带来了极大的便利,是用户常用的功能。由于微信

收稿日期:2019-12-06;**修回日期**:2020-02-21;**录用日期**:2020-02-26。　**基金项目**:国家重点研发计划项目(2016YFE0206700)。
作者简介:梁登高(1995—),男,贵州兴仁人,硕士研究生,主要研究方向:网络流量分析、恶意流量分析; 周安民(1963—),男,四川成都人,研究员,主要研究方向:安全防御管理、移动互联网安全; 郑荣锋(1990—),男,重庆人,博士研究生,主要研究方向:信息系统安全; 刘亮(1982—),男,四川叙永人,高级工程师,博士研究生,主要研究方向:系统安全、恶意代码分析、漏洞挖掘与利用、网络应用安全; 丁建伟(1986—),男,四川自贡人,高级工程师,博士,主要研究方向:网络威胁情报、暗网数据分析。

具有庞大的用户量,且涉及资金业务,微信红包与转账功能往往成为不法分子的攻击目标,如针对性的钓鱼链接、虚假的红包图案。“拼手气红包”的出现,无疑增添了微信中群成员的活跃度,然而不法者却利用该功能聚集人员在线赌博,借助私人转账与红包功能进行赌博结算。除此之外,微信中红包与转账功能也有被作为非法交易的在线资金转账的可能性。

基于以上所述,对微信中红包与转账行为的准确识别具有极大的现实意义:1)准确识别出正常的红包与转账行为,可避免点击虚假红包链接;2)实时检测一段时间范围内红包与转账行为发生的次数,可用于发现赌博行为;3)线上检测红包与转账行为可作为相关部门的一种取证手段。当前,针对微信中用户行为识别的研究极少,文献[3]中对微信中通信协议进行了详细研究,该研究将微信中用户行为分为七种类别,其中将红包、转账、支付归为支付功能一类,并没有对红包与转账功能进行细化识别。文献[4]中基于机器学习对微信中文本与图片信息进行分类,但该方法仅针对文本与图片信息。文献[5]中将微信中传输文本、图片、红包和转账功能细化为16种用户行为,其中红包与转账功能共被细化为14种操作,故识别一次发送或接收红包行为需要结合多种操作的识别结果,这使得最终的识别准确率较低。文献[6]中针对移动手机App的加密流量,提出了一种对服务类别进行分类的方法与系统,选取了微信与WhatsApp[7]作为测试对象,该方法在数据预处理上结合了层次聚类、隐马尔可夫模型(Hidden Markov Model, HMM)和阈值启发式等方法,时间开销相对大,且最终的准确率相对较低。

文献[8]中“斯诺登事件”表明现已有在线拦截和窃听用户的行为,这涉及了对用户行为识别的大量研究。贾军等[9]提出一种基于深度包检测(Deep Packet Inspection,DPI)自关联的分类方法来对应用类别进行识别,准确率低。文献[10]提出了UIPicker,一种识别用户敏感信息输入的智能框架,即通过自动检测用户输入,从而识别是否为敏感信息。文献[11]通过提取数据包IP、TCP报头信息,从而建模识别用户的具体应用行为,该方法不采集应用负载信息,故针对特征集中于负载的应用识别率不高。与文献[11]相似,文献[12]也只采集IP、TCP报头信息,开发了AppScanner框架,通过提取54种网络流双向特征,从而对应用进行建模与识别,但该框架不对应用内具体行为进行识别。文献[13]中通过分析基于超文本传输安全(HyperText Transfer Protocol over Secure Socket Layer, HTTPS)协议加密的手机应用流量,从而识别用户行为,研究中以Gmail、Facebook和Twitter三种应用作为实验对象。文献[14]中提出一种基于层次聚类的有监督机器学习方法,用于识别移动用户的在线行为分类,研究中以Kakao Talk[15]为实验对象。文献[16]中选取谷歌应用商店中110种APP作为实验对象,提出了一种程序指纹生成的方法。文献[17]的研究提出了一种智能手机应用识别的系统,但仅选取了13种应用作为实验对象。以上的方法,多数实验选取的样例较少,一些方法得到的准确率也相对较低。

与现有研究不同,本文提出的是一种新的用户行为识别方法。将用户的一个单任务,如微信中“发送红包”,映射为一个大的流量块,而组成“发送红包”的一系列单操作,如“点击红包按钮”“输入金额”“输入密码”分别映射为小的流量块。如此划分使得小流量块可表征每个特定单操作的特征,而这些小流量块的顺序则表征着整个大流量块的特征,即从多个维度刻画用户单任务的特征,最终得到综合特征。本文研究工作的贡献在于提出了一种新的用户行为识别方法,并能准确识别出微信中红包与转账行为以及这些行为发生的次数。

1　微信中红包与转账功能的特征分析

1.1　协议层面

微信中多数通信行为采用了专有协议MMTLS,其开发者称该协议是在传输层安全协议版本1.3(Transport Layer Security (TLS) Protocol Version 1.3,TLS1.3)的基础上做的更改[18]。红包与转账行为只采用了基于MMTLS的超文本传输协议(HyperText Transfer Protocol, HTTP),即在传输过程中通过MMTLS对应用层数据进行封装,然后通过普通的HTTP协议将其传输到目标地址。结合文献[3]中的研究工作,将微信中基于MMTLS的HTTP协议进行通信的服务类别汇总如表1(“✓”代表相应服务使用了相应协议,否则代表无)。

表1　基于MMTLS的HTTP协议进行通信的服务汇总

Tab. 1　Service summary of communication for MMTLS-based HTTP

服务	MMTLS-based HTTP	服务	MMTLS-based HTTP
浏览朋友圈	✓	广告点击	✓
支付操作	✓	聊天	
点击小程序	✓	文件传输	
浏览订阅号	✓		

通过Wireshark[19]抓包与分析发现,基于MMTLS的HTTP有两个显著的特征:其一是在发送过程中,开头都会冠以“POST/mmtls/”字样[3];其二是服务端的端口都是80。这两点将有助于收集数据时,准确打标签。

1.2　用户操作步骤细化

移动应用中的单个功能多由数次用户操作来完成,如“发送文件”需先找寻文件路径,接着点击发送。显然,微信中收发红包与转账也如此。本文实验的研究目的是识别用户在某一时间内发送或接收了红包与转账的次数,故发送与接收被示为相互独立的行为。研究中要识别四种行为,即“发送红包”“接收红包”“发起转账”“接收转账”。这四种行为分别由多个单操作组成,见表2。要特别说明,表2对四种行为的细化仅为表述目标行为的细节特征,从而为后续提出的大、小流量分块机制提供物理依据。

表2　微信中收发红包与转账功能操作细化说明

Tab. 2　Operations of sending and receiving red packets and fund transfer in WeChat

操作顺序	行为			
	发送红包	接收红包	发起转账	接收转账
1	点击红包	点击图标	点击转账	点击图标
2	填写金额	打开红包	填写金额	确认收款
3	输入指纹或密码	…	输入指纹或密码	…

2　本文方法

2.1　收集数据

为了保证数据集标签的准确性,不同测试机以及不同行为的数据将被单独收集。以“发送红包”为例,仅将测试机连入无线网卡,通过Wireshark抓取该无线网卡的数据。四种目标行为皆是由基于MMTLS协议的HTTP来传输数据,利用Wireshark中的过滤功能可将测试机中相应操作的数据准确抓取。微信中其他功能(见表1)也采用了基于MMTLS协议的HTTP来传输数据,但在进行红包以及转账操作时,不可能会

同步产生该类功能的数据，这保证了数据标签的准确性。为避免手机操作系统与账号对实验结果造成影响，在实验设备上采用了IOS（Iphone5s、Iphone5）和Android（小米5s、小米5x、谷歌手机LG E960 Nexus 4）手机，同时采用了多个账号来进行实验。微信中其他功能产生的数据（特别是使用基于MMTLS协议的HTTP进行通信）以及在收集数据过程中其他APP产生的数据都将被归为一个类，在实验中标为“普通数据”。另外，目标行为使用TCP，收集过程将过滤掉非TCP的数据。

2.2 突发流量分块

2.2.1 大突发与小突发

在移动终端，当用户与远端进行交互操作，如单击一个功能按钮，都会产生一系列的数据包。这些数据包的各项属性由相应通信协议严格控制，不仅有多种相同属性，如源IP、目的IP、协议、目的端口等，且严格按协议定义的次序发送，即时间间隔与协议、网络状态相关。可将这些数据包归为一个流量突发块，由时间阈值来划分。本实验中将单操作产生的流量块称作小突发块，用来划分的阈值称为小突发阈值。

单个功能任务通常由多个操作组成，因这些连续操作由人来实施，故单个操作之间存在明显的时间间隔。在正常情况下，这一时间间隔是在一定范围内的（除非网络故障、设备故障等）。这一时间间隔由用户的操作速度、网络状态与应用的反应时间决定。人类的操作速度与计算机的操作速度不在一个级别，且加上应用反应时间与网络传输时间，故单个操作之间的时间间隔必定大于小突发阈值，这一特性是对单个功能的流量划分大、小突发块的物理基础。而同一功能任务的单操作之间按严格的次序发生，相邻单操作之间的时间间隔有上限（正常情况下），设置特定的阈值可将不同单任务之间的数据包分隔开。本实验中将该时间阈值称为大突发阈值，即时间间隔在大突发阈值之内的相邻数据包属于同一个大突发块，显然一个大突发块将会包含多个小突发块。

不同的服务类别发送的服务端地址不同，在突发流量的划分中加入IP地址的限定，可最大限度把同一时间不同服务的数据界定开。本实验中，大、小突发流量块的划分都是基于两个IP地址之间的通信数据。在文献[5]和文献[16]的研究工作中皆提到了对“突发流量块”的定义，即接收时间间隔在突发阈值之内的所有网络数据包（对源或目的地地址不做要求）属于同一个突发块。在本文的研究中对突发块的定义不同，定义如下：

定义1 在两个通信IP地址之间，时间间隔在小突发阈值T_S之内的相邻数据包，属于同一个小突发块。

定义2 在两个通信IP地址之间，时间间隔在大突发阈值T_L之内的相邻小突发块，属于同一个大突发块。具体的划分如图1所示。

2.2.2 突发阈值

将流量数据划分为大、小突发块，对应了手机应用中单个任务由多个用户操作组成的特性，即小突发映射单个小操作，而大突发映射单个任务。两个层次的划分由对应的两个突发阈值来决定。文献[20]中的研究表明，智能手机中95%的数据包在前一个数据包的4.5 s内被接收或传输。文献[16]中的研究将1 s作为突发阈值，文献[5]中采用的是1.25 s。这些研究中的突发阈值对应的是本研究中的小突发阈值，它们没有大突发的概念。

大突发阈值的确定极其重要，它划分单任务间的界限，是最终识别用户“发送红包”“接收红包”“发起转账”与“接收转账”次数的最大依据。大突发阈值与用户的操作习惯有一定关系，但除了特殊情况，如用户中断任务、网络卡顿、设备损坏等，不同用户之间的操作速度出入不大。在收集数据时，通过多人使用习惯性的速度来操作手机。

大突发阈值将单任务产生的流量框定之后，利用单操作流量的突发性再次将大流量块细分，可对单任务的特征做进一步剖析。小突发阈值用于准确划分单操作产生的数据包，它只与对应的通信协议以及网络传输有关。本研究中详细的大小阈值验证实验见3.2节。

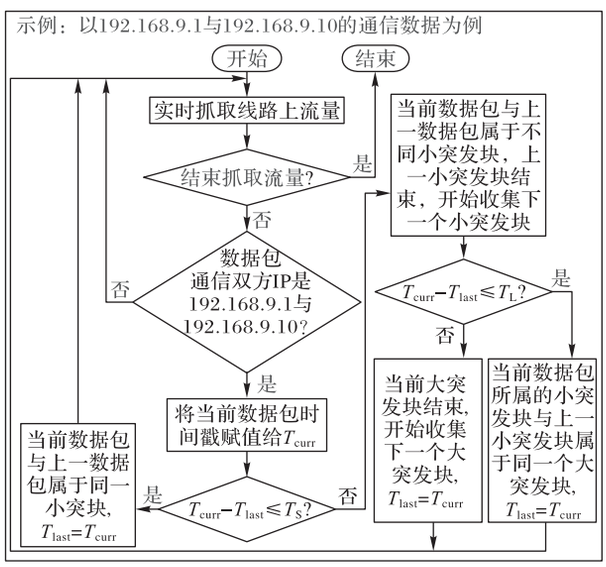

图1 突发流量数据分块示意图

Fig. 1 Blocking diagram of burst traffic

2.3 特征提取

2.3.1 相关定义

在详细讲解特征提取之前，先对本实验中用到的概念作定义和说明。

定义3 一定时间内具有相同五元组{源IP，源端口，目的IP，目的端口，协议}的数据包属于一个单向流，即上行流或下行流。其中发起连接的一方发出的数据包为上行流。同一个通信会话的上行流与下行流的组合称为流（flow），定义如下。

上行流：

$$\{src_{ip}=ip_1, des_{ip}=ip_2, src_{port}=port_1, des_{port}=port_2, \text{TCP}\} \tag{1}$$

下行流：

$$\{src_{ip}=ip_2, des_{ip}=ip_1, src_{port}=port_2, des_{port}=port_1, \text{TCP}\} \tag{2}$$

流（TCP）：

$$\{ip_1, port_1 \Leftrightarrow ip_2, port_2\} \tag{3}$$

特别说明：

1）在实验中，对流进行组合时，限于同一个小突发块，即不跨越小突发块组流。

2）完整流。在小突发块里，一个TCP流，若其不包含建立连接的三次握手以及释放连接的四次握手，则将其标记为不完整流。

2.3.2 多层次特征

接下来将分别从流、小突发块、大突发块三个层次详细讲

解本实验提取特征的过程,最终用于训练的是大突发块的综合特征。

流特征　本实验将对小突发块中的单个流提取12个统计特征,分别是数据包总量、带负载的数据包总数、上行流的带负载数据包总数量、下行流的带负载数据包总数、流持续时间、流数据包之间的平均时间间隔、上行流数据包之间平均时间间隔、下行流数据包之间平均时间间隔、上行流的源端口、上行流数据包的平均负载长度、下行流数据包的平均负载长度、是否为完整流。

小突发层　小突发块映射一个用户单次操作,不同的操作产生的流量有很大差异,如流数量、数据包平均长度等。对小突发块中的每个流提取以上提到的12个特征之后,通过综合计算得到小突发层的特征。其中除了"上行流的源端口"与"是否为完整流"两个特征的处理方法不同之外,其余10个特征皆是取平均值。"上行流的源端口"将被用来计算相邻(按时间顺序)流的上行源端口间的差值的绝对值,并取多流间该值的平均值。"是否为完整流"被用来统计小突发中不完整的流的总数。除了以上特征,小突发块将会新加一个特征,该特征是流的总数,故对单个小突发块将提取13个特征。

大突发层　本实验将一个任务映射为一个大突发块,不同的任务由不同的用户单操作组成。故大突发块内的小突发块数量是一个极其重要的特征。通过分析发现,微信中红包与转账功能的小突发数量差异性不大,但它们小突发中的流的数量有一些细微差别,故为了实现对红包与转账功能的准确区分,每个大突发块中的前三个小突发块的流的数量将作为三个重要特征。大突发块中所有小突发的流的数量总和能很好地刻画大突发的总体特征。除了以上新增的5个特征,余下的特征将从大突发块中各个小突发块的特征中获取。其中小突发中"流的总数"与"不完整流的总数"将通过求和的方式产生大突发块的流总数与不完整的流总数,而其余11个特征将通过求平均值的方法产生大突发块的11个特征,故总的综合特征为18个特征,见表3。

表3　实验特征

Tab. 3　Experimental features

特征标识	特征含义	特征标识	特征含义
bur_num	小突发总数	avg_num_up_dt	流的上行带负载的平均数据包数
num_flow_bur1	第一个小突发包含的流总数	avg_num_dw_dt	流的下行带负载的平均数据包数
num_flow_bur2	第二个小突发包含的流总数	avg_duratio	流的平均持续时间
num_flow_bur3	第三个小突发包含的流总数	avg_f_tm	流的数据包间平均时间间隔
num_flows	包含的流总数	avg_f_up_tm	上行流的数据包间平均时间间隔
num_tl_flows	完整的流总数	avg_f_dw_tm	下行流的数据包间平均时间间隔
num_ntl_flows	不完整的流总数	avg_spt_int	按时间顺序的相邻的流上行源端口的绝对平均差值
avg_num_pack	流的平均数据包总数	avg_up_bt	上行流数据包平均字节数
avg_num_p_dt	流中带负载的平均数据包数	avg_dw_bt	下行流数据包平均字节数

通过XGBoost(eXtreme Gradient Boosting)算法训练得出这些特征的重要性如图2,结果显示,除了"不完整流总数"重要性极低之外,其余特征对分类准确率贡献度都很高。

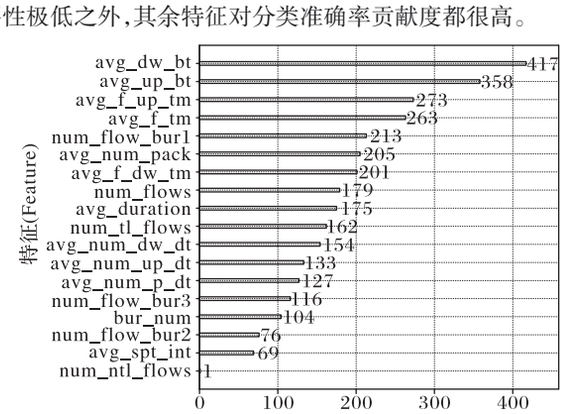

图2　特征重要性(具体特征含义见表3)

Fig. 2　Feature importance (specific feature meanings are shown in Tab. 3)

2.4　训练分类器

特征提取完成之后,得到了带标签的数据集。将数据划分为测试集与训练集,对数据进行建模。

3　方法评估

3.1　数据收集

按照2.1节描述的方法收集实验数据,不同设备和账号的样本量保持平衡,并准确对其进行打标签,收集的数据汇总如表4。

表4　实验数据

Tab. 4　Experimental data

系统	数据描述	发送红包	接收红包	发起转账	接收转账	其他
Android	单任务	130	130	130	130	
	数据包	8 868	2 734	8 807	4 130	60 928
IOS	单任务	130	130	130	130	
	数据包	9 272	2 946	11 224	6 036	33 276

3.2　阈值验证实验

突发阈值与网络环境以及应用程序内部开发逻辑相关(详见2.2.1节),故需要对具体网络环境的大量样本进行验证实验,才能得出具体环境的最优值。本实验收集的实验数据如表4所示。

在阈值验证实验中,初步将小突发阈值定为1 s(结合已有研究与经验),大突发阈值的取值范围定为5~14 s(实际操作时估测的范围),同时记录不同取值下单任务流量的划分准确率(根据数据标签,假设某一"发送红包"样本包含10个数据包,根据对应阈值对该样本的划分情况计算被正确划分的数据包所占比例)与分类器(使用XGBoost)准确率。结合两个准确率的结果,得出表现最佳的大突发阈值为11.8 s,如图3所示。

大突发阈值确定之后,将大突发阈值设为11.8 s,小突发阈值取值范围定为1.0~1.6 s,同时分别训练分类器,最终实验结果见图4。从图4中的结果可知,小突发阈值为1.26 s时,得到的分类器准确率最高。需要特别说明的是,阈值验证

实验与3.3节中的对照实验独立进行，得到的最高准确率存在些许偏差。

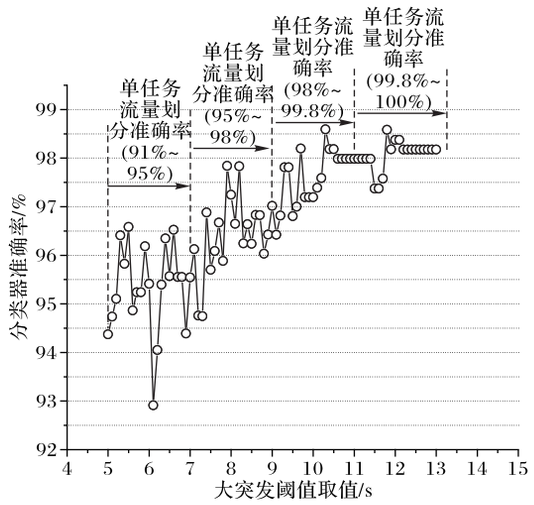

图3 大突发阈值取不同值时准确率的变化

Fig. 3 Accuracy varying with different large burst thresholds

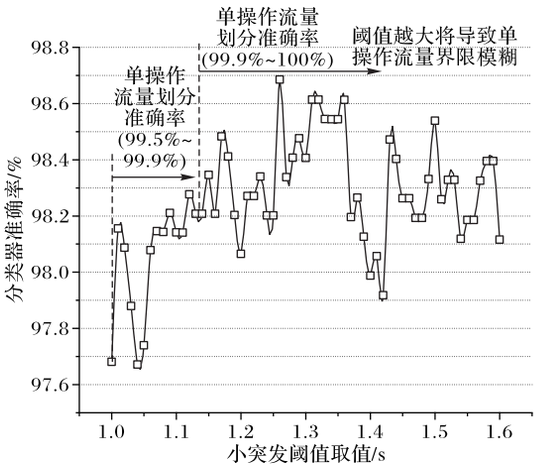

图4 小突发阈值取不同值时准确率的变化

Fig. 4 Accuracy varying with different small burst thresholds

3.3 对照实验

为了验证所提方法的有效性，将从两方面入手。首先是选用5种经典的机器学习算法来验证所提方法的算法普适性，算法分别是随机森林(Random Forest，RF)、支持向量机(Support Vector Machine，SVM)、*K*最近邻(*K*-Nearest Neighbor，*K*NN)、C4.5决策树和XGBoost，所有算法采用默认参数(每种算法都可对参数进一步调优，从而得到更好的识别率。实验中皆采用默认参数，旨在验证所提方法可应用于多个已有算法，而不仅仅只针对具体的算法。实验中采用的默认参数为weka[21]的推荐参数，已进行过初步优化，感兴趣的读者可对具体方法进行深度的优化)，训练集均采用十折交叉验证，每个实验重复10次，结果取平均值。第二个方面是为了验证所提方法的识别率，将与现有研究工作进行对照实验。选择了文献[5]与文献[3]中的研究方法，文献[5]中通过划分突发流量块来识别表2中的每个单操作，进而识别红包与转账的行为与次数，其研究目标与本实验相同。文献[3]对微信中通信协议进行研究，基于流的层面来提取特征和识别行为，当前多数研究工作都是基于流的层面来识别行为。最终的实验结果如表5。

表5 三种方法准确率对比 单位：%

Tab. 5 Accuracy comparison of three methods Unit: %

方法	RF	C4.5	*K*NN	SVM	XGB
本文方法	97.5	96.5	95.2	84.3	97.6
文献[5]方法	95.4	94.6	94.3	38.5	95.6
文献[3]方法	94.3	94.5	95.1	41.8	95.4

从表5可看出，所提出的方法在五种算法下，除了SVM为84.3%，其余准确率皆高于95%，其中RF与XGBoost的准确率高达97.5%和97.6%，这表明所提出方法具有强普适性。三种方法的实验对比结果显示，所提出的方法在五种算法下准确率皆为最高，且在SVM的表现远高于另两种方法。需要特别提出的是：表中文献[3]方法的准确率是对每个流样本的识别准确率，即该方法只能识别一条流是否为目标行为的数据，而一次目标行为包含很多流，这意味着，该方法无法识别出目标行为的次数；表中文献[5]方法的准确率是单个操作的平均准确率，而目标行为包含多个单操作(见表2)。

表6是三种方法对相同的目标行为进行提取特征所需的平均时间与所产生的平均样本量。从表中可看出，基于单个流提取特征(文献[3]方法)的时间与平均产生的样本量都高于其余两种基于流量块的方法，这是可预见的结果。单任务由多种用户单操作组成，故而会产生多个流。这也进一步表明，仅基于流来识别由多次单操作组成的用户行为，将会需要更大的时间和空间上的开销。所提出的方法在时间和产生的平均样本上皆低于文献[5]中的方法，这也是可预见的结果。所提方法将每次单任务所产生的流量看作一个整体，得出的识别率是针对单任务。文献[5]中将单任务切割为多个不相关的单操作(仅有顺序之分)，分别计算特征与识别，这也意味着多个单操作的识别误差将对最终的单任务识别误差产生乘法扩散。

综合以上结果，所提方法在时间、空间以及识别率上皆具有突出的表现，在多种算法下保持高准确率。其中在XGBoost下表现最好，对其将进行进一步的实验评估。

表6 三种方法单次任务提取特征的效率对比

Tab. 6 Efficiency comparison of three methods for single task feature extraction

方法	发送红包		接收红包		发起转账		接收转账	
	平均时间/ms	样本量	平均时间/ms	样本量	平均时间/ms	样本量	平均时间/ms	样本量
本文方法	54	1	41	1	56	1	43	1
文献[5]方法	69	3	44	2	74	3	48	2
文献[3]方法	90	7	43	2	90	7	61	3

3.4 XGBoost建模评估

3.4.1 评估指标

本实验中用到的评估指标分别是准确率(Accuracy)、召回率(Recall)和受试者工作特征曲线(Receiver Operating Characteristic，ROC)曲线。在给出这些指标的详细定义前，先对一些相关变量进行定义。

1) *TP*(True Positive)：正样本被正确分类的数量，即真阳性。

2)*FP*(False Positive):负样本被错误分类的数量,即假阳性。

3)*TN*(True Negatives):负样本被正确分类的数量,即真阴性。

4)*FN*(False Negatives):正样本被错误分类的数量,即假阴性。

准确率是指被正确分类的样本占总样本的比例,体现的是分类器的预测性能,其越大代表分类效果越佳,计算如下:

$$Accuracy_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{4}$$

其中i代表多分类中的具体类别。

召回率是指被正确分类的正样本数占正样本总数的比例,体现的是分类器对正样本的预测性能,即查全率。计算如下:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{5}$$

ROC曲线是一条横坐标为假正率(False Positive Rate,FPR),纵坐标为真正率(True Positive Rate,TPR)的曲线。假正率与真正率的相关定义见式(6)、(7)。ROC曲线可直观地观察分类器的准确性,其形状越向左上方扩展则表征分类性能越好,曲线与横坐标的围成面积(Area Under the Curve,AUC)可用来表征分类准确性。

假正率计算公式:

$$FPR_i = \frac{FP_i}{TN_i + FP_i} \tag{6}$$

真正率计算公式:

$$TPR_i = \frac{TP_i}{TP_i + FP_i} \tag{7}$$

3.4.2　评估结果

用于评估的算法为XGBoost,测评得出的混淆矩阵如图5。混淆矩阵的结果表明,“发送红包”与“发起转账”的测试漏报率为零,而“接收转账”与“普通数据”漏报率也只是2%,只有“接收红包”的漏报率为7%,表明所提出的方法的分类效果极佳。

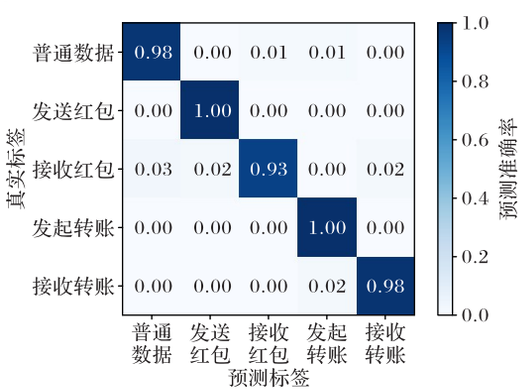

图5　混淆矩阵

Fig. 5　Confusion matrix

由混淆矩阵的结果分别计算准确率与召回率,并绘制ROC曲线进一步地评估所提的方法的性能。从表7可看出,平均准确率为97.58%,平均召回率为97.72%,“发送红包”与“发起转账”的召回率高达100%。

图6为预测结果的ROC曲线,其中“发送红包”与“发起转账”的AUC值达到了1,而“普通数据”与“接收转账”也为0.98及以上,相对较低的“接收红包”也高达0.96,总平均值为0.98。

表7　评估结果　　单位:%

Tab. 7　Results of evaluation　　unit: %

行为	准确率	召回率	行为	准确率	召回率
普通数据	97.6	97.6	发起转账	96.5	100.0
发送红包	97.8	100.0	接收转账	97.8	97.8
接收红包	98.2	93.2			

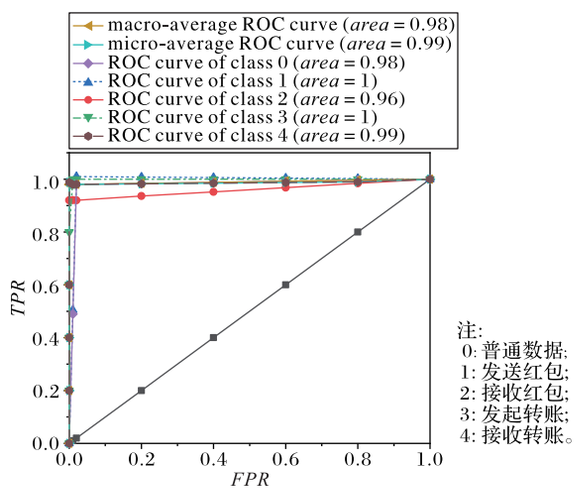

图6　预测结果ROC曲线

Fig. 6　ROC curve of prediction result

3.4.3　场景测评

本文研究工作要实现的目标不仅是从加密的流量中识别“发送红包”“接收红包”“发起转账”与“接收转账”四种行为,并且要在用户连续操作产生的流量中,识别出这四种行为发生的次数。故要将真实环境中新收集的流量(不加任何处理与过滤手段,不打标签)来对训练好的模型进行真实场景测评,测评数据与结果见表8。表中c_1~c_4分别代表“发送红包”“接收红包”“发起转账”与“接收转账”。

表8　真实场景测试数据以及测评结果

Tab. 8　Testing data and evaluation results in real senses

数据集	时间/min	样本数量	包含行为次数				预测到行为次数			
			c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
Normal	60	417	0	0	0	0	1	1	5	5
Data1	4	18	5	5	0	0	5	5	0	0
Data2	4	12	1	1	5	5	1	1	5	5
Data3	60	174	50	50	0	0	50	52	0	4
Data4	60	196	0	0	50	50	0	5	50	53
Data5	60	255	25	25	25	25	26	23	32	24

从两个方面分析表8中的结果。首先,Normal的结果显示“普通数据”的漏报率为2.88%(Normal中总样本数为417,不包含四种目标行为,有1+1+5+5=12个样本漏报,即被预判为其他四种行为,漏报率约为12/417=2.88%),这表明所提方法可将97.12%的非四种目标行为的数据排除开来。另一方面从时间上来看,Data1与Data2是短时间内连续操作多次目标行为,结果显示识别率为100%。Data3与Data4是长时间内分别对收发红包与转账功能进行测试,结果显示误报率极低。Data5中包含了所有行为的数据,从检测结果上来看,除了“发起转账”的误报率稍微高一点之外(Data5中含有25次“发起

转账”行为，却预测到32次，误报数为|32-25|=7，其余三种分别为1，2，1），其他目标行为的检测率都符合预期。

4 结语

准确识别微信中收发红包与转账行为有助于辅助相关部门对相关违法活动（如红包赌博、非正常资金来往等）取证，以及排除虚假红包与转账链接。针对现有研究在识别微信支付行为上效率以及准确率低的情况，本文提出了一种识别微信中收发红包与转账行为的方法，将一次用户单任务（如发送红包）行为映射为一定时间范围内（大突发阈值）的数据块，提取特征时将数据块再细化为多个小数据块，这些小数据块映射的是组成单任务行为的多个用户操作（如“点击红包按钮”“输入金额”“输入密码或指纹”），从而提炼出单任务行为的多维度特征。在实验环境下，经多种经典机器学习与对照实验的验证，结果显示所提出的方法在时间效率、空间占用率、识别准确率、算法普适性等方面皆具有优越的性能，最高的准确率可达97.58%。在真实场景的测试下，所提方法性能优越，基本实现了在连续的四种目标行为数据中识别出对应行为的次数的研究目标。

显然，划分大、小突发流量块的机制针对的是由多种单操作组成的应用行为的识别，而现今移动应用行为几乎皆属此类，故接下来可将此方法的应用场景扩大化。

参考文献（References）

[1] 腾讯科技. 2018微信年度数据报告[EB/OL]. [2019-01-09]. https://mp. weixin. qq. com/s/Khf4-dChUIgIg8NHCiJPyg. (Tencent. 2018 WeChat annual report[EB/OL]. [2019-01-09]. https://mp. weixin. qq. com/s/Khf4-dChUIgIg8NHCiJPyg.

[2] 腾讯科技. 微信发布2019春节数据报告：8.23亿人收发红包[EB/OL]. [2019-02-12]. https://mp. weixin. qq. com/s/aBV5PP1h6Mh0zMqPJNZiPw. (Tencent. WeChat 2019 Spring festival data report: 8.23 million people send and receive red packets [EB/OL]. [2019-02-12]. https://mp. weixin. qq. com/s/Khf4-dChUIgIg8NHCiJPyg.

[3] HOU C, SHI J, KANG C, et al. Classifying user activities in the encrypted WeChat traffic[C]// Proceedings of the IEEE 37th International Performance Computing and Communications Conference. Piscataway: IEEE, 2018:1-8.

[4] SHAFIQ M, YU X, LOGHARI A A, et al. WeChat text and picture messages service flow traffic classification using machine learning technique[C]// Proceedings of the IEEE 18th International Conference on High Performance Computing and Communications/ IEEE 14th International Conference on Smart City/ IEEE 2nd International Conference on Data Science and Systems. Piscataway: IEEE, 2016: 58-62.

[5] YAN F, XU M, QIAO T, et al. Identifying WeChat red packets and fund transfers via analyzing encrypted network traffic[C]// Proceedings of the 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/ 12th IEEE International Conference on Big Data Science and Engineering. Piscataway: IEEE, 2018:1426-1432.

[6] FU Y, XIONG H, LU X, et al. Service usage classification with encrypted internet traffic in mobile messaging apps[J]. IEEE Transactions on Mobile Computing, 2016, 15(11):2851-2864.

[7] WhatsApp. WhatsApp Web[EB/OL]. [2019-09-18]. https://web.whatsapp. com.

[8] LANDAU S. Making sense from Snowden: what's significant in the NSA surveillance revelations [J]. IEEE Security and Privacy, 2013, 11(4):54-63.

[9] 贾军，杨进，李涛. 一种基于DPI自关联数据包检测分类方法[J]. 四川大学学报（自然科学版），2019，56(1)：29-36.（JIA J, YANG J, LI T. A DPI-based autocorrelation method for packet detection classification[J]. Journal of Sichuan University (Natural Science Edition), 2019, 56(1):29-36.）

[10] NAN Y, YANG Z, YANG M, et al. Identifying user-input privacy in mobile applications at a large scale[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(3): 647-661.

[11] ALAN H F, KAUR J. Can Android applications be identified using only TCP/IP headers of their launch time traffic[C]// Proceedings of the 9th ACM Conference on Security and Privacy in Wireless and Mobile Networks. New York: ACM, 2016: 61-66.

[12] TAYOR V F, SPOLAOR R, CONTI M, et al. AppScanner: automatic fingerprinting of smartphone apps from encrypted network traffic[C]// Proceedings of the 2016 IEEE European Symposium on Security and Privacy. Piscataway: IEEE, 2016: 439-454.

[13] CONTI M, MANCINI L V, SPOLAOR R, et al. Can't you hear me knocking: identification of user actions on Android apps via traffic analysis[C]// Proceedings of the 5th ACM Conference on Data and Application Security and Privacy. New York: ACM, 2015:297-304.

[14] PARK K, KIM H. Encryption is not enough: inferring user activities on Kakao Talk with traffic analysis [C]// Proceedings of the16th International Workshop on Information Security Applications, LNCS 9503. Cham: Springer, 2015:254-265.

[15] Kakao. Kakao Web[EB/OL]. [2019-09-18]. https://www. kakaocorp. com.

[16] TAYLOR V F, SPOLAOR R, CONTI M, et al. Robust smartphone app identification via encrypted network traffic analysis[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(1):63-78.

[17] WANG Q, YAHYAVI A, KEMME B, et al. I know what you did on your smartphone: inferring app usage over encrypted data traffic [C]// Proceedings of the 2015 IEEE Conference on Communications and Network Security. Piscataway: IEEE, 2015:433-441.

[18] Tencent. MMTLS: introduction of TLSV1.3 based Tencent security communication protocol [EB/OL]. [2019-09-18]. https://github. com/WeMobileDev/article/blob/master/SUMMARY. md.

[19] Wireshark. Wireshark Web[EB/OL]. [2019-01-08]. https://www.wireshark. org/.

[20] CONTI M, MANCINI L, SPOLAOR R, et al. Analyzing Android encrypted network traffic to identify user actions[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(1): 114-125.

[21] Weka. Weka3[EB/OL]. [2019-01-08]. https://www. cs. waikato. ac. nz/ml/weka//.

This work is partially supported by National Key Research and Development Program of China (2016YFE0206700).

LIANG Denggao, born in 1995, M. S. candidate. His research interests include network traffic analysis, malicious traffic analysis.

ZHOU Anmin, born in 1963, research fellow. His research interests include security defense and management, mobile internet security.

ZHENG Rongfeng, born in 1990, Ph. D. candidate. His research interests include information system security.

LIU Liang, born in 1982, senior engineer, Ph. D. candidate.. His research interests include system security, malicious code analysis, vulnerability mining and utilization, network application security.

DING Jianwei, born in 1986, Ph. D., senior engineer. His research interests include cyber threat intelligence, dark Web data analysis.