一种基于机器学习的 Tor 网络识别探测技术

张 玲[1],卫传征[1],林臻彪[1],段琳琳[2]
(1.北京赛博兴安科技有限公司,北京 102200;2.郑州大学 信息工程学院,河南 郑州 450001)

摘 要:Tor 是一种基于洋葱路由通信协议建立的隐蔽加密通信系统。该系统基于互联网现有路由、数据加密等协议,构建了一套保护通信实体的身份隐匿机制,使得经过 Tor 网络传播的数据难以被有效追踪和分析。然而近年来这项隐蔽通信技术被罪犯大量使用,已成为网络犯罪和非法交易的温床。为有效应对该问题,提出一项基于机器学习的Tor 网络识别检测技术,通过主动生成 Tor 网络流量,基于机器学习技术实施流特征提取与检测,从而发现参与 Tor 通信的网络实体及其通信类型,进而检出潜在的恶意暗网用户。实验表明,该方法可有效识别 Tor 通信实体以及通信行为,如电子邮件和 FTP 应用等。

关键词:暗网探测;Tor;通信实体识别;机器学习
中图分类号:TN918 **文献标识码**:A DOI:10.16157/j.issn.0258-7998.200759

中文引用格式:张玲,卫传征,林臻彪,等. 一种基于机器学习的 Tor 网络识别探测技术[J].电子技术应用,2021,47(4):54-58.
英文引用格式:Zhang Ling,Wei Chuanzheng,Lin Zhenbiao,et al. A method for identifying Tor hosts based on machine learning techniques[J]. Application of Electronic Technique,2021,47(4):54-58.

A method for identifying Tor hosts based on machine learning techniques

Zhang Ling[1],Wei Chuanzheng[1],Lin Zhenbiao[1],Duan Linlin[2]
(1.Beijing Cyber XingAn Technology Co.,Ltd.,Beijing 102200,China;
2.School of Information Engineering,Zhengzhou University,Zhengzhou 450001,China)

Abstract:Tor is an anonymous Internet communication system based on onion routing network protocol. Network traffics generated by normal applications become hard to trace when they are delivered by Tor system. However, an increasing number of cyber criminals are utilizing Tor to remain anonymous while carrying out their crimes or make illegal transactions. As a countermeasure, this paper presents a method able to identify Tor traffics and thereby recognize related Tor hosts. The method proposes several groups of features extracted from network traffic and resort to machine learning algorithm to evaluate feature effectiveness. Experiments in real world dataset demonstrate that the proposed method is able to distinguish Tor flows from normal traffics as well as recognize the kind of activity in Tor generated by different normal applications.

Key words:darknet detection;Tor;communication entity recognition;machine learning

0 引言

Tor 匿名网络是一个由全球志愿者维护的各自匿名网络所组成的大型分布式匿名通信网络,其核心技术是美国海军研究室开发的洋葱路由系统,设计初衷是保护政府机关的数据通信隐私。

Tor 用户通过连接一系列虚拟通道在通信的源端与目的端之间建立间接的数据链路,使得包括个人和机构在内的用户在互联网中的数据传输行为匿名化[1]。由于该技术能够有效规避网络监管,成为访问受限网段的有效措施。

洋葱路由技术提供的身份匿名性和数据安全性使得 Tor 网络成为网络内容犯罪的温床。同时,区块链、虚拟数字货币等技术的发展为网上非法交易带来便利,更使得包括 Tor 网络在内的暗网成为互联网中的法外之地,产生越来越多涉及黄、暴、恐的非法信息和非法交易。鉴于此,本文研究 Tor 网络流量的分析和识别。对于给定的真实网络数据,本研究的目标是鉴别其中流量是通过普通网络通信数据还是 Tor 流量。在有效识别 Tor 流量基础上,本文进一步研究 Tor 通信行为分类,包括浏览网页、邮件服务、即时通信、流媒体、FTP、VoIP 和 P2P 通信等。

1 相关工作

(1)Tor 流量识别

近年来,研究人员提出了若干解决方案来识别 Tor 网络中产生的数据[2]。LASHKARI A H 等人[3]对网络中 Tor 客户端与 Tor 网络入口节点之间的通信进行时序分

析，通过随机森林分类器对提取的特征达成95%的识别准确率。

HE G等人[4]提取Tor流量特征，通过隐马尔科夫模型将Tor通信分离为P2P、FTP、即时通信和网页4类，并取得了92%的准确率。

CHAKRAVARTY S等人[5]提出了一个针对Tor客户端的流量攻击技术来识别Tor客户端的IP地址。对Tor服务端节点发起主动流量分析攻击，并观测客户端侧产生的摄动，通过统计相关性指标可以识别到一组Tor流量相关的服务端和客户端。

(2)Tor网站识别

针对Tor网站的攻击包括网站指纹识别攻击和洋葱站点枚举攻击[6]。最早针对Tor网站的指纹攻击由HERRMANN D等人[7]提出，并只取得了3%的正确率。该研究领域接下来获得了极大的重视和发展，并在不同的场景环境下取得了超过90%准确率，相关研究见参考文献[8]、[9]。

访问Tor网站极大地依赖洋葱网络中的隐藏服务目录查询协议[10]，而该协议被证明容易被仅具有低带宽Tor中继节点的攻击者进行服务枚举攻击[11]，从而测量Tor网站等洋葱服务的活跃程度。已有相关工作通过枚举攻击方法来研究洋葱服务的生态[12]。当然，该类攻击的威胁可能随着相关协议的更新得到缓解。

上述研究工作主要关注两类问题：(1)有效识别互联网中Tor节点之间传输的流量；(2)对于Tor通信流量所承载的应用数据，识别其应用服务类型。与上述研究类似，本文通过机器学习方法来识别网络中的Tor流量，并进一步识别Tor流量中承载的8类应用数据。实验表明，本文提出的方法能够取得较高的预测准确率。

2 本文方法

本文首先通过设置网络探针捕获网络中的流量数据，进而对流量中混杂的数据帧进行组流，将属于相同数据流的帧按照协议和帧顺序进行恢复。然后以数据流为样本，提取其机器学习分类特征，训练分类器，并应用训练后的分类器对目标数据进行分类，分析分类效果。

需要特别解释的是，本文中一条通信流由一组具有相同五元组{源IP、目的IP、源端口、目的端口、协议}的数据帧组成，其中Tor支持的协议为TCP协议。本文依照该规则，在TCP协议层进行流重组。

2.1 数据采集

数据准备阶段包括数据生成、数据采集和组流3个环节。首先需要生成带有类别标签的数据：在沙箱中分别运行包括网页浏览、即时通信、音频流、视频流、电子邮件、VoIP、P2P数据传输和FTP文件传输8种类型的网络应用，并利用架设的Tor网关服务将相应应用产生的网络流打包传输至Tor网络中。进而使用tcpdump应用在Tor网关两端采集数据帧，即可得到两类实验数据：(1)应用于Tor流量检测的带Tor和非Tor标签的网络帧数据；(2)应用于Tor通信行为分类的带8种应用类型标签的Tor网络帧数据。

接下来，按照五元组将网络帧进行重组，形成网络流。对具有相同五元组的不同网络流，采用TCP协议中的FIN帧来进行切分。8类应用数据流的生成方法描述如下：

(1)页面浏览：通过selenium自动化工具调用Firefox和Chrome浏览器的geckodriver和chromium内核，遍历访问Alexa知名网站列表，并对首页内超链接进行深度为2的访问遍历，获得所有HTTP和HTTPS流量，数据总量8.5 GB。

(2)即时通信：本文采集的即时通信数据来自微信、QQ、Skype、Telegram、WhatsApp、和Signal，行为包括文本聊天、文件传输等，数据总量1.9 GB。

(3)音频流：QQ音乐和网易云音乐是中国最大的音频流媒体应用平台，本文分别采集这两个桌面应用自动播放时产生的4.0 GB流媒体传输数据。

(4)视频流：本文采集了腾讯视频、搜狐视频、优酷视频等自动播放时的7.8 GB多媒体流。

(5)电子邮件：通过邮件客户端绑定包括采用STMP/S、POP3/SSL和IMAP/SSL等协议的网络邮箱，除通过邮件客户端的自动更新功能进行邮件传输外，主动通过各个邮箱发送、接收邮件及其附件，数据总量2.3 GB。

(6)VoIP：采集包括微信语音、Skype通话、Facebook Messenger通话和Google Voice等在内的语音通话数据作为VoIP标签数据，数据总量2.0 GB。

(7)P2P：国内最知名的P2P应用是迅雷，然而由于其广告等扩展功能太多，为获得干净的P2P流量，本实验采用Bittorrent应用进行数据传输，获得26 GB P2P协议数据。

(8)FTP：本文在采集FileZilla的客户端和服务端应用进行文件上传和下载时产生的流量作为FTP标签数据，采集的数据包括SFTP协议数据和FTPS协议数据，数据总量10 GB。

2.2 特征提取

本文对同一个网络流中的上下行流量分别提取特征(规定客户端指向服务端的方向为上行方向，其反方向为下行方向)，最终形成27个的特征：

(1)上行帧时间差(UPward Inter Arrival Time，UIAT)：上行帧之间的时间差，包括时间差的均值F_0、最小值F_1、最大值F_2和标准差F_3。

(2)下行帧时间差(Downward Inter Arrival Time，DIAT)：下行帧之间的时间差，包括均值F_4、最小值F_5、最大值F_6和标准差F_7。

(3)帧时间差(Flow Inter Arrival Time，FlowIAT)：所有帧之间的时间差，包括其均值F_8、最小值F_9、最大值F_{10}和

标准差 F_{11}。

(4)流活跃时间(Active):在流进入空闲状态之前所经历的时间,包括其均值 F_{12}、最小值 F_{13}、最大值 F_{14} 和标准差 F_{15}。

(5)流空闲时间(Idle):在流进行活跃状态之前,保持空闲状态的时长,包括其均值 F_{16}、最小值 F_{17}、最大值 F_{18} 和标准差 F_{19}。

(6)流字节速率(Flow Bytes Per Second,FlowBPS):该流平均每秒传输的字节数,用 F_{20} 表示。

(7)流帧率(Flow Packets Per Second,FlowPPS):该流平均每秒传输的帧数量,用 F_{21} 表示。

(8)流负载(Flow Payloads,FP):该流上行字节数 F_{22}、下行字节数 F_{23}、上行帧数 F_{24}、下行帧数 F_{25}。

(9)流持续时间(duration):流的第一帧和最后一帧的间隔时间 F_{26}。

上述特征除流持续时间外,按照特点可分为 6 组特征。前 3 组分别是上行帧时间差、下行帧时间差和帧时间差特征,着重刻画上下行流量中的时间间隔特征,分别命名为 UIAT、DIAT 和 FlowIAT;第 4、5 组特征关注流在活跃和空间状态之间变化的特点,分别用 Active 和 Idle 指代;最后一组包括流字节速率 FlowBPS、流帧率 FlowPPS 和流负载 FP,关注的是流在不同层面的传输量和传输速率,统一用 FP 代替。

2.3 实验流程

首先通过实验验证所提特征集合对 Tor 流量和普通流量进行区分的能力。实验分两个阶段,包括:(1)通过假设检验,验证每个特征在 Tor 流量和普通流量上数值分布上的差异显著性;(2)通过训练分类器,验证提取的特征对 Tor 流量和普通流量进行分类的有效性。

在假设检验阶段中,设零假设 H_0 为:对于 Tor 网络流和正常流提取的特征,不存在统计上的显著差异性。进而采用 SPSS 软件中非参检验工具集,分别进行 Mann-Whitney (MW,$p=0.05$) 测试和 Kolmogorov-Smirnov (KS,$p=0.05$)测试[13],以增强结论的可靠性。相关测试均可用于验证目标数据与给定分布之间的差异性,实验中风险阈值设定为 0.05。

第二阶段,采用机器学习分类器进行效果评估,即基于本文所提的 27 维分类特征,采用 10 折交叉验证,分别在 Tor 流量检测问题和 Tor 通信行为分类中测试分类器的效果。

本文采用 scikit-learn 机器学习工具集[14],从中分别选择 K 近邻分类器(K Nearest Neighbor,KNN)[15]、逻辑回归分类器(Logistic Regression,LR)、C4.5 决策树(Decision Tree,DT)[16]、朴素贝叶斯分类器(Naïve Bayesian,NB)[17]、支撑向量机分类器(Support Vector Machine,SVM)[18]和随机森林分类器(Random Forest,RF)[19]6 种经典分类模型进行分类效果测评。

3 实验结果

3.1 评价指标

本实验分别采用准确率(Precision)、召回率(Recall)、F-测度(F-Measure)、马修斯相关系数(Matthews Correlation Coefficient,MCC)[20]、接受者工作特征曲线(ROC)和精度-召回率曲线(PRC)等测量指标评价 Tor 流量的分类效果。

3.2 实验结果

本节分别进行 Tor 流量识别和 Tor 通信行为分类。在 Tor 流量识别任务中,首先通过假设检验分别验证所提特征在不同类别流量下的分布差异性,进而通过训练分类器验证特征的有效性;在 Tor 通信行为分类中,直接通过训练分类器验证特征的有效性。

3.2.1 Tor 流量识别

经过 MW 和 KS 检验,不同特征的参数分布均拒绝原假设 H_0,故备择假设成立,表明本文所提特征在 Tor 流量和普通流量上的概率分布呈现显著差异性,可被有效用于区别两类数据流,因而运用所有 27 个特征来完成接下来的分类任务。

表 1 给出了 Tor 流量识别问题的实验结果,可以看到,所有给出的分类算法都能够有效识别 Tor 网络流量。其中,随机森林分类器的预测结果最好,在每个评价指标上都取得了最高分;支撑向量机的分类效果次之;效果最差的分类器是决策树,在 F-Measure 等 4 个综合评价指标中都得到了最低的得分。

表 1 全特征 Tor 流量识别

分类器	Precision	Recall	F-Measure	MCC	ROC	PRC
K 近邻	0.801	0.836	0.818	0.829	0.820	0.827
逻辑回归	0.891	0.898	0.894	0.893	0.900	0.898
决策树	0.718	0.780	0.748	0.738	0.799	0.801
朴素贝叶斯	0.723	0.885	0.795	0.811	0.827	0.885
支撑向量机	0.982	0.982	0.982	0.978	0.979	0.979
随机森林	0.984	0.983	0.984	0.939	0.998	0.998

表 2 展示了不同特征组在Tor 流量识别任务中的效果,分类算法采用全特征预测时效果最好的支撑向量机和随机森林。由表可知,在两个分类器中,特征组 Active 和特征组 Idle 的预测效果均最差,仅略高于随机分类,特征组 FP 均取得了最好的预测效果;UIAT、DIAT 和 FlowIAT 的分类效果接近,预测效果介于上述两类特征的预测效果之间。

接下来通过逻辑回归分类器分析预测特征。逻辑回归分类器的优点是除能够给出分类预测结果外,还可以检验自变量与因变量的相关性,并且与自变量相对应的回归系数可以显示自变量与因变量的相关强度以及正负相关性。本文用逻辑回归分类器判断基于内容特征的各项指标是否具有较强的链路预测能力。

对于该二分类问题,将某条流量的类别作为因变量;

表 2 特征组 Tor 流量识别

分类器	特征组	Precision	Recall	F-Measure	MCC	ROC	PRC
支撑向量机	UIAT	0.730	0.718	0.769	0.759	0.756	0.761
	DIAT	0.665	0.713	0.688	0.691	0.687	0.633
	FlowIAT	0.716	0.791	0.752	0.748	0.750	0.757
	Active	0.570	0.595	0.582	0.593	0.581	0.582
	Idle	0.601	0.591	0.596	0.597	0.593	0.596
	FP	0.936	0.944	0.940	0.944	0.942	0.944
随机森林	UIAT	0.689	0.734	0.711	0.709	0.701	0.720
	DIAT	0.747	0.633	0.685	0.691	0.687	0.684
	FlowIAT	0.691	0.690	0.690	0.690	0.691	0.691
	Active	0.543	0.588	0.565	0.571	0.569	0.569
	Idle	0.612	0.598	0.605	0.611	0.603	0.606
	FP	0.910	0.930	0.920	0.927	0.927	0.928

将 27 维特征作为自变量，可得到表 3 逻辑回归实验结果的训练结果。其中标准误差项用于评价回归系数是否显著不为 0，z 值由回归系数与标准误差之比得到，与 p 一起用来检验回归系数为 0 的零假设。此处同样取显著性水平 α 为 0.05，当 p 小于 α 时，说明回归系数显著不为 0，即自变量与因变量显著相关。由 p 列可知，特征 F_{19} 的 p 值大于显著性水平 α，因此除特征“流空闲时间的标准差值”外，在回归模型中，其他预测特征对 Tor 流量识别均具有显著影响。

表 3 逻辑回归实验结果

特征	回归系数	标准误差	z	p
F_0	-3.26	2.78	-1.17	0.000
F_1	1 283.12	87.30	14.69	0.000
F_2	57.45	10.67	5.38	0.000
F_3	921.48	60.36	15.26	0.000
F_4	-324.82	23.17	-14.00	0.000
F_5	68.51	3.47	19.71	0.001
F_6	131.23	20.78	6.33	0.000
F_7	-143.91	27.38	-5.25	0.000
F_8	66.22	4.87	13.57	0.000
F_9	-149.54	28.97	-5.16	0.000
F_{10}	-474.69	39.99	-11.8	0.000
F_{11}	150.58	12.38	12.15	0.000
F_{12}	190.39	14.28	13.32	0.000
F_{13}	699.31	32.87	21.26	0.000
F_{14}	-98.03	18.89	-5.22	0.002
F_{15}	-185.07	97.37	-1.90	0.000
F_{16}	410.25	21.32	19.23	0.000
F_{17}	63.89	7.67	8.30	0.000
F_{18}	920.91	38.98	23.62	0.000
F_{19}	8.60	2.38	3.60	0.199
F_{20}	22.87	2.98	7.64	0.000
F_{21}	-475.91	19.98	-23.82	0.000
F_{22}	425.37	23.84	17.83	0.000
F_{23}	580.61	43.28	13.41	0.000
F_{24}	38.26	8.87	4.31	0.000
F_{25}	1.87×10^6	9.99×10^4	18.89	0.000
F_{26}	-2.14×10^5	7 219.00	-29.79	0.000

3.2.2 Tor 通信行为分类

表 4 展示了本文所提特征在 Tor 通信行为分类任务中的结果。由表可知，支撑向量机分类器取得了最好的预测效果，综合指标值接近 0.85。相较于支撑向量机和随机森林，K 近邻、决策树和逻辑回归的模型拟合能力较差，因而预测效果较差。

4 结论

本文针对 Tor 流量的检测与识别问题，提出了一种基于机器学习的 Tor 流量探测技术。通过主动生成 Tor 网络流量，提取网络流特征，训练分类器，对 Tor 流量与普通流量进行分类，取得了 F-measure 测度值 0.98 的效果。更进一步地，在对 Tor 服务所承载的多种应用数据进行分类时，该套特征也取得了准确率 85.9%和召回率 82.1%的预测结果。试验中支撑向量机和随机森林在所有实验条件下取得了最好的分类效果，效果最好的分类特征组是 FP。实验表明本文所提方法是有效的。

表 4 全特征 Tor 通信行为分类

分类器	Precision	Recall	F-Measure	MCC	ROC	PRC
K 近邻	0.712	0.609	0.656	0.650	0.652	0.651
逻辑回归	0.698	0.693	0.695	0.634	0.703	0.697
决策树	0.581	0.630	0.605	0.598	0.610	0.603
朴素贝叶斯	0.648	0.599	0.623	0.627	0.631	0.634
支撑向量机	0.859	0.821	0.840	0.850	0.848	0.843
随机森林	0.813	0.786	0.799	0.807	0.803	0.806

本文的后续工作将考虑在时序特征、网络拓扑结构特征等方面进一步丰富和优化特征集合，以进一步提高 Tor 通信类型分类的效果。同时进一步研究 Tor 网络中的其他常见通信数据，拓展本文所提方法的应用范围。

参考文献

[1] MCCOY D, BAUER K, GRUNWALD D, et al. Shining light in dark places: understanding the Tor network[C]. Privacy Enhancing Technologies, PETs 2008, 2008: 63-76.

[2] BASYONI L, FETAIS N, ERBAD A, et al. Traffic analysis attacks on Tor: a survey[C]. 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies(ICIoT), Doha, Qatar, 2020: 183-188.

[3] LASHKARI A H, DRAPER G G, MAMUN M S I, et al. Characterization of Tor traffic using time based features[C]. International Conference on Information Systems Security, 2017: 253-262.

[4] HE G, YANG M, LUO J, et al. Inferring application type information from Tor encrypted traffic[C]. International Conference on Advanced Cloud and Big Data, 2014: 220-227.

[5] CHAKRAVARTY S, BARBERA M V, PORTOKALIDIS G, et al.

《电子技术应用》http://www.chinaaet.com

On the effectiveness of traffic analysis against anonymity networks using flow records[C].International Conference on Passive and Active Network Measurement,2014:247-257.
[6] JANSEN R,JUAREZ M,GALVEZ R,et al.Inside job:applying traffic analysis to measure Tor from within[C].NDSS,2018.
[7] HERRMANN D,WENDOLSKY R,FEDERRATH H,et al.Website fingerprinting:attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier[C].IEEE International Conference on Cloud Computing Technology and Science,2009:31-42.
[8] HAYES J,DANESIZ G.K-fingerprinting:a robust scalable website fingerprinting technique[C].Usenix Security Symposium,2016:1187-1203.
[9] PANCHENKO A,LANZE F,ZINNEN A,et al.Website fingerprinting at internet scale[C].Network and Distributed System Security Symposium,2016.
[10] NEPAL S,DAHAL S,SHIN S,et al.Deanonymizing schemes of hidden services in tor network:a survey[C].International Conference on Information Networking.IEEE,2015.
[11] BIRYUKOV A,PUSTOGAROV I,WEINMANN R P.Trawling for Tor hidden services:detection,measurement,deanonymization[C].2013 IEEE Symposium on Security and Privacy,2013:80-94.
[12] BIRYUKOV A,PUSTOGAROV I,THILL F,et al.Content and popularity analysis of Tor hidden services[C].2014 IEEE 34th International Conference on Distributed Computing Systems Workshops,2014:188-193.
[13] ÖZÇOMAK M S,KARTAL M,SENGER O,et al.Comparison of the powers of the kolmogorov-smirnov two-sample test and the mann-whitney test for different kurtosis and skewness coefficients using the monte carlo simulation method[J].Journal of Statal & Econometric Methods,2013,2(4):81-98.
[14] PEDREGOSA F,VAROQUAUX G,GRAMFORT A,et al.Scikit-learn:machine learning in python[J].Journal of Machine Learning Research,2011:2825-2830.
[15] CUNNINGHAM P,DELANY S J.K-nearest neighbour classifiers[Z].2007.
[16] PENG J,LEE K,INGERSOLL G.An introduction to logistic regression analysis and reporting[J].Journal of Educational Research,2012:3-14.
[17] KAVIANI P,DHOTRE S.Short survey on naive bayes algorithm[J].International Journal of Advance Research in Computer Science and Management,2017.
[18] Tian Yingjie,Shi Yong,Liu Xiaohui.Recent advances on support vector machines research[J].Technological and Economic Development of Economy,2012,18(1):5-33.
[19] ALI J,KHAN R,AHMAD N,et al.Random forests and decision trees[J].International Journal of Computer Science Issues,2012.
[20] CHICCO D,JURMAN G.The advantages of the Matthews correlation coefficient(MCC) over F1 score and accuracy in binary classification evaluation[J].BMC Genomics,2020.

(收稿日期:2020-07-16)

作者简介:
张玲(1976-),女,博士研究生,高级工程师,主要研究方向:网络安全、数据分析。
卫传征(1984-),男,硕士研究生,工程师,主要研究方向:网络安全。
段琳琳(1974-),女,博士研究生,讲师,主要研究方向:数据分析与信号处理。

(上接第 53 页)

system[C].Proceedings of the 19th ACM Conference on Computer and Communications Security,2012.
[40] WANG Q,GONG X,NGUYEN C T K,et al.CensorSpoofer:asymmetric communication using IP spoofling for the censorship-resistant Web browsing[C].Proceedings of the 19th ACM Conference on Computer and Communications Security,2012.
[41] KADIANAKIS G.Pluggable-transports/obfsproxy obfs2[DB/OL].[2020-10-12].https://gitweb.Torprojet.Org/pluggable-transports/obfsproxy.Git/tree/doc/obfs2/obfs2-protocol-spec.txt.
[42] KADIANAKIS G.Pluggable-transports/obfsproxy obfs2[DB/OL].[2020-10-12].https://gitweb.Torprojet.Org/pluggable-transports/obfsproxy.Git/tree/doc/obfs3/obfs3-protocol-spec.txt.
[43] ANGEL Y.Obfs4/blog/master/doc/obfs4-spec.txt obfs4[DB/OL].[2020-10-12].https://github.com/Yawning/obfs4/blog/master/doc/obfs4-spec.txt.
[44] FIFILED D,LAN C,HYNES R,et al.Block-resistant communication through domain fronting[C].Proceedings on Privacy Enhancing Technologies,2015.
[45] DYER K P,COULL S E,RISTENPART T,et al.Protocol misidentification made easy with format-transforming encryption[C].Proceedings of ACM SIGSAC Conference on Computer and Communications Security,2013.

(收稿日期:2020-10-12)

作者简介:
陈欢(1995-),男,硕士研究生,主要研究方向:匿名通信、暗网探测。
苏马婧(1985-),女,博士,正高级工程师,主要研究方向:网络空间测绘、网络安全。
王学宾(1986-),男,博士,工程师,主要研究方向:计算机网络、信息安全等。

版权声明