Journal of Computer Applications
计算机应用, 2020, 40(6): 1685 - 1691
ISSN 1001-9081
CODEN JYIIDU
2020-06-10
http://www.joca.cn

文章编号:1001-9081(2020)06-1685-07
DOI:10.11772/j.issn.1001-9081.2019111981

基于图像纹理的网站指纹技术

张道维*,段海新
(清华大学 网络与信息安全实验室,北京 100084)
(*通信作者电子邮箱 zhangdaowei2018@gmail.com)

摘　要:网站指纹技术能够让本地监听者通过审查用户与Tor入口节点之间的匿名流量从而追踪到该用户访问的具体网站。现有的研究方法只提取了匿名流量中的部分元数据来进行网站指纹的刻画,忽视了大量隐含的指纹信息。为此,提出了基于图像纹理和深度卷积神经网络的网站指纹技术Image-FP。首先,将匿名通信流量映射成RGB彩色图;然后,使用残差神经网络(ResNet)构造出能进行自主特征学习的网站指纹分类模型。在50个网站构成的封闭世界场景下,Image-FP能够取得97.2%的分类准确率,相较于最前沿的网站指纹攻击技术提高了0.4个百分点。而在更接近真实环境的开放世界场景中,Image-FP能够以100%的准确率识别出监控网站的流量,其准确性和鲁棒性更是远远高于其他指纹技术。实验结果表明,匿名流量图像化的技术能够更多地保留网站指纹的相关特征,并且在避免复杂特征工程的同时,能够进一步提高分类精度。

关键词:匿名网络;Tor;网站指纹;数据可视化;卷积神经网络
中图分类号:TP393　　**文献标志码**:A

Website fingerprinting technique based on image texture

ZHANG Daowei*, DUAN Haixin
(*Network and Information Security Lab*, *Tsinghua University*, *Beijing* 100084, *China*)

Abstract: Website fingerprinting technique enables the local monitor to track which websites a user is visiting by capturing anonymous traffic between that user and the Tor (The onion router) entry nodes. Prior researches only extract part meta-data in the anonymous traffic to construct website fingerprints, and ignore much hidden fingerprint information inside the traffic. Therefore, a website fingerprinting technique named Image FingerPrinting (Image-FP) and based on deep convolutional neural network and image texture was proposed. Firstly, the anonymous communication traffic was mapped into Red-Green-Blue (RGB) images. Then, the Residual Network (ResNet) was used to construct the website fingerprinting model with automatic feature learning ability. In a closed-world scenario of 50 websites, Image-FP obtained classification accuracy of 97.2%, which is 0.4 percentage points higher than that of the state-of-the-art website fingerprinting attack technique. In the open-world scenario which is more realistic, Image-FP can identify the traffic of monitored websites with 100% accuracy, has the strongest accuracy and robustness among all fingerprinting techniques. The experimental results demonstrate that, the technique of converting anonymous traffic into images can preserve more features relevant to the website fingerprints, and further improve the classification accuracy while avoiding complex feature engineering

Key words: anonymous network; The onion router (Tor); website fingerprinting; data visualization; Convolutional Neural Network (CNN)

0　引言

Tor(The onion router)是当前规模最大、使用最广泛的低延迟匿名通信网络。数百万的用户每天通过Tor来匿名访问网站,从而来隐蔽自己的网络活动。Tor一方面使用多层的传输层安全(Transport Layer Security, TLS)协议进行数据加密;另一方面使用重路由技术进行路由转发策略来达到匿名的目的。由于Tor匿名网络中存在大量非法活动,因此对于Tor的去匿名化研究一直是国内外的研究热点。早期的研究方向主要集中在针对Tor匿名流量的识别上。何高峰等[1]在2013年提出了基于TLS报文长度分布的识别方法,并取得了90%以上的准确率。为了应对此类网络审查,各种隐蔽的接入技术相继被提出,其中Meek通道技术因其具有较高的隐蔽性而被广泛使用。何永忠等[2]在此基础上提出了针对Meek混淆技术的识别方法,同样能够取得90%的识别准确率。

近些年来的研究表明,还存在一系列侧信道攻击技术使得攻击者能够通过分析Tor的匿名流量从而推断出用户具体的网络行为,其中最具代表性的攻击方式为网站指纹攻击(Website Fingerprinting Attack),其核心思想在于:对于不同的网站,其网页内容都大不相同(如网页代码、图片、脚本、样式表等),因此尽管通信内容加密,浏览器在加载网页时产生的匿名流量元数据也不尽相同。攻击者首先通过对每一个网站都建立独特的指纹模型,并根据这些指纹特征训练出合适的

收稿日期:2019-11-21;**修回日期**:2019-12-18;**录用日期**:2019-12-24。
作者简介:张道维(1993—),男,台湾台北人,硕士研究生,主要研究方向:网络安全、深度学习;　段海新(1972—),男,山东济宁人,教授,博士,主要研究方向:网络安全、网络测量。

分类器;之后便可以采集用户的访问流量,并使用该分类器进行流量的分类,从而定位出用户访问的具体是哪些网站。

在网站指纹攻击的场景下,攻击者只能对网络流量进行本地、被动的监听。这里"本地"表示的是攻击者控制了用户与Tor入口节点之间链路上的任意节点,包括路由器、局域网管理权限、互联网服务提供商(Internet Service Provider, ISP)、自治系统(Autonomous System, AS)甚至可以是恶意的入口节点;"被动"指的是攻击者可以记录用户所有的通信流量,但是不能篡改、延迟或丢弃任何数据包。当然攻击者也无法解密通信流量,否则就能通过数据包内容轻易得知用户访问的目标站点了。

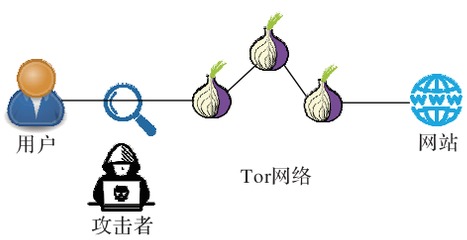


图1 网站指纹攻击威胁模型

Fig. 1 Threat model of website fingerprinting attack

网站指纹攻击通常在两种场景下进行评估,即封闭世界(closed-world)和开放世界(open-world)。在封闭世界场景下,用户被限定只能访问有限集合里的网站。通常网站集合规模较小,攻击者有能力对每个网站都收集足够的流量数据进行网站指纹分类器的训练。然而一些研究[3-4]认为该场景是不符合实际情况的,因为攻击者无法确认用户可能访问的网站集合,同时当网站集合规模较大时,为每一个网站都建立单独的指纹特征模型是不现实的。因此后续的研究中又提出了更加贴近真实环境的开放世界。在开放世界场景下,用户可以访问任意网站,而攻击者的目的是判断用户访问的网站是否在自己的监控列表范围内。

在2009年,Herrmann等[5]首次将网站指纹技术应用在Tor匿名网络中,将数据包长度的频率分布作为网站指纹特征并使用朴素贝叶斯作为分类器,然而最终准确率只有3%。Panchenko等[6]在相同数据集上,通过引入流量的独特突发性(burstiness)这一新特征并结合支持向量机(Support Vector Machine, SVM)进行网站指纹分类,将准确率提升至55%。Wang等[7]提取了超过3 000维特征向量来进行网站指纹建模,并使用基于加权的距离度量指标和K最近邻(K-Nearest Neighbor, KNN)分类器来衡量网站指纹的相似度,最终取得91%的准确率。

在此之后,各种网站指纹技术相继被提出并保持较高的分类准确率。本文将着重介绍其中4种效果最出色的先进方法,并在后续实验中重新进行评估。

1)CUMUL(CUMULation)。Panchenko等[8]基于累积包大小这一新特征提出了CUMUL技术。该方法将通信流量视作时序序列,其中出口方向数据包大小为正数,而入口方向数据包大小为负数。累积包特征的第1个值是该时序序列第1个数据包的大小,而第i个值是坐标$i-1$的值与第i个数据包大小的和,以此类推。由于CUMUL使用以高斯径向基函数为核函数的SVM分类器,需要输入数据的特征维度保持相同,因此CUMUL通过对累积包特征插值100个点从而得到100个特征值。再加上入口/出口方向的数据包总数量和总字节大小,最终共有104个维度的特征。在100个网站构成的封闭世界中,CUMUL达到91%的准确率;而在网站数量为9 000的开放世界中,CUMUL方法的真阳性率(True Positive Rate, TPR)为96%,假阳性率(False Positive Rate, FPR)为1.98%。

2)k-FP(k-FingerPrinting)。Hayes等[9]从高达4 000维的特征中选取了其中最重要的150个作为描述网站指纹的特征,并结合随机森林分类器进行网站指纹攻击。在55个网站构成的封闭世界中,k-FP可以获得91%的准确率;而对于开放世界问题,k-FP将随机森林的叶子节点构造成新特征作为KNN分类器的输入,最终得到88%的TPR和0.5%的FPR。

3)AWF(Automated Website Fingerprinting)。Rimmer等[10]首次将深度学习方法应用在网站指纹攻击中。AWF将网站的每条访问实例都表示成由±1组成的时序序列,其中的符号表示数据包的方向,即出口方向的数据包为+1,而入口方向的数据包为-1。同时使用卷积神经网络(Convolutional Neural Network, CNN)作为分类器进行网站指纹攻击。AWF利用深度学习自主特征学习的优势,避免了如CUMUL和k-FP方法所需的特征工程步骤。在100个网站构成的封闭世界中,并且在每个网站都具有2 500条访问实例时,AWF可以达到96.3%的分类准确率;而在开放世界场景下,AWF可以获得70.9%的TPR和3.8%的FPR。

4)DF(Deep Fingerprinting)。Sirinam等[11]在AWF的基础上设计了更复杂的CNN模型,该网络具有更深的结构同时引入了更多的卷积层以及批规范化(Batch Normalization, BN)。DF使用与AWF相同的±1时序序列作为模型输入,在由95个网站构成的封闭世界中,DF能够取得98.3%的分类准确率,高于CUMUL、k-FP和AWF这3种方法;而在世界大小为20 000的开放世界中,DF能够取得95.7%的TPR和0.7%的FPR。

以上研究表明,深度学习在网站指纹领域中具有极大的潜力。深度学习算法无需对输入数据进行繁杂的特征工程,而能够自主地从大量原始数据中挖掘相关的抽象特征。相较于传统机器学习算法,基于深度学习的网站指纹模型具有更高的分类准确率。然而文献[10-11]仅使用"数据包方向"作为卷积神经网络模型的输入,相较于原始的匿名流量数据无疑丢失了部分信息。

在此基础上,本文提出了全新的基于图像纹理和深度卷积神经网络的网站指纹技术Image-FP(Image FingerPrinting)。Image-FP通过将匿名通信流量映射成RGB(Red-Green-Blue)彩色图的形式,从而保留了最完整的原始信息。之后再利用残差网络(Residual Networks, ResNet)在图像分类上的优势,构造出能够自主特征学习的网站指纹分类模型。相较于过往的网站指纹技术,Image-FP在封闭世界和开放世界场景下均取得了最优的效果。

1 数据收集

1.1 收集方法

使用10台基于OpenStack的虚拟云服务器作为匿名流量采集的设备,其中每台服务器分别拥有1个CPU和2 GB的内存空间。使用tor-browser-crawler开源程序来驱动Tor浏览器(版本8.5.4)访问网站。相较于wget或curl等工具,使用Tor浏览器能够更加真实地模拟用户的浏览行为。此外,Tor浏览器同时也是Tor官方推荐的接入Tor网络的方法,在具备简易

的操作性的同时,也保证高度的匿名性。

使用轮替策略(round-robin)来进行数据的采集。假设封闭世界集合中包含50个网站,则在每一批次的流程中,都会对每一个网站连续访问25次,之后再进行下一网站的访问,总共进行5个批次的采集。每个网站都有300 s的时间去完全加载其页面上的内容,若加载完成后会在该页面停留10 s。之后便会关闭Tor浏览器进程并清除所有配置文件信息。使用tcpdump程序去捕获每个Tor进程的流量数据,同时过滤掉那些不属于用户与Tor入口节点之间的通信流量。本文会保留每条访问实例的完整数据包,也即是说,每条访问实例都会以1个pcap包进行表示。整个数据集的采集流程持续2星期。

此外,对Tor程序的配置文件进行了部分改动。将其中的MaxCircuitDirtiness字段从原来的600 s增加至600 000 s,使得通信链路不会每10 min就进行重新构建,从而保证了数据采集的稳定性。在每一批次的流程结束后,会关闭旧的链路并重新建立新的链路,避免对该链路节点造成过大的负载压力。另外,将UseEntryGuard字段设置为0,从而避免在数据采集过程中只使用3个固定的入口节点,同时也可以给数据集带来更多样的变化,增加分类器模型的泛化性。

1.2　数据集

由于没有权威的数据显示Tor用户访问哪些网站的频率更高,而且对于敏感网站的定义每个地区也都不同,同时本文的目的主要在于比较不同网站指纹攻击方法的性能优劣,因此对于具体网站的选择并不会有太大影响。基于上述原因,选择Alexa服务中的网站作为数据集中的网站列表。Alexa网站排名根据流量数据来评估网站的受欢迎度,在学术界中使用非常广泛,同时也多次应用在与Tor的相关研究中。

1.2.1　封闭世界

选择Alexa排名前50的网站来组成封闭世界数据集。每台设备都会访问各网站125次,因此每个网站都会有1 250条访问实例,即数据集中总共包含62 500条流量数据。而在具体域名的选择上,遵从以下3个要点:

1)对于域名相同而仅仅是顶级域不同的网站,只保留其中一个。例如Google搜索页面根据地理位置(国家)的不同存在许多不同的站点:google. com(通用)、google. de(德国)、google. co. jp(日本)等。这些网站在内容上相似,而且如果审查者的目的是封锁Google搜索引擎的流量,也无需具体区分这些网站。

2)对于上述二级域重复的网站,选择保留使用国家顶级域的网站,即google. de而不是google. com。由于Google搜索网站会根据用户地理位置的不同跳转到相对应的版本,也就是说,如果处在德国的用户访问google. com就会被重新导向至google. de。而Tor在建立匿名链路时的出口节点是随机的,所以如果使用google. com会导致实际上每次访问的网站不同,导致网站内容的不一致性。使用固定国家顶级域的网址则可以避免此现象。

3)对于域名不同但是网站内容相同的网址,也只保留其中一个,如blogger. com和blogspot. com。

1.2.2　开放世界

开放世界的网站集合由监控网站和非监控网站两部分组成。对于监控网站的部分,使用封闭世界的数据集。

而在非监控网站的部分,选择Alexa排名前60 000的网站,并去掉封闭世界数据集中包含的网址。使用相同的10台设备进行流量数据的采集,但是与封闭世界方法不同的是:由于这些非监控网站是作为背景流量,因此对于这些网站只进行1次访问。每台设备都会按照顺序依次访问6 000个网站,同时对每个网站页面都进行快照。通过将那些页面空白、访问失败或超时错误的访问实例移除后,最终获得50 000条非监控网站的流量数据集。因此,开放世界数据集中一共包含112 500条访问实例。

2　Image-FP网站指纹技术

2.1　数据表示

将网站的每1个访问实例都保存为1个pcap包。由于这些pcap包中的流量都是加密的,因此不再试图从这些加密流量中提取指纹特征,而是进一步直接将pcap包视作二进制文件。

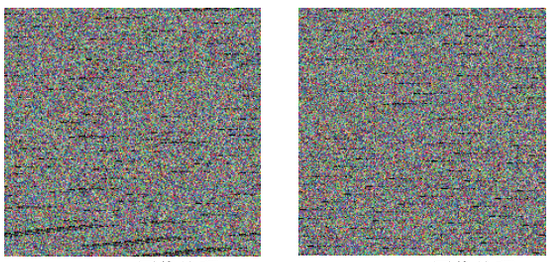
对于二进制文件,其中的每个字节范围都在00和FF之间,刚好对应灰度图的1个像素点(灰度图的像素取值范围为0~255,其中0为黑色,255为白色)。因此二进制文件可以按此方法映射为1张灰度图。这种二进制文件图像化的方法首次由Nataraj等[12]提出,通过该技术将恶意代码样本以灰度图的形式展现出来,并利用图像中的纹理特征对恶意代码家族进行聚类。

在此基础上,Image-FP使用类似的图像化技术,进一步将pcap包映射成RGB彩色图的形式。对于二进制文件,其中连续的3个字节都可以分别对应于RGB图像中的3个通道,即第1个字节对应R通道的值,第2个字节对应于G通道的值,而第3个字节对应于B通道的值。换句话说,每3个字节都对应RGB图像中的一个像素,若最末端数据不足3个像素,则不足的部分用FF值进行填充。假设有1个十六进制串为05 B1 29 43 CA 7E 9F FD,那么按照每3个字节为1组,可以得到(05,B1,29),(43,CA,7E),(9F,FD,FF),再使用如下公式将每一组值转换为R、G、B的3个通道的值:

$$\begin{cases} R = \sum_{i=0}^{1} h_i \times 16^i \\ G = \sum_{i=0}^{1} h_{i+2} \times 16^i \\ B = \sum_{i=0}^{1} h_{i+4} \times 16^i \end{cases}$$

因此最终每个像素点为(5,177,29),(43,202,126),(159,253,255)。之后,将得到的像素点按照三维矩阵"长度×宽度×3"的方式进行排列,便可以得到1张RGB彩色图。这里需要特别说明的是,由于数据集中pcap包大小普遍大于1 000 KB,为了能够最大限度地保留所有信息,因此选择使用RGB图像而不是灰度图像,同时将长度和宽度都设定为1 024个像素。此外,会去除pcap包头前434个字节,因为这部分是所有pcap包含的相同头部数据。

图2展示了Google和Youtube两个不同网站的访问实例图像化后的RGB彩色图。由图2可以看出,两者的图像纹理有些许的不同,同时Google样本的图像底部有3条明显的黑色条纹,表明两者之间的图像纹理确实存在差异性,因此可以将这种图像纹理作为网站指纹特征进一步去进行分类判断。



(a) Google图像纹理 (b) Youtube图像纹理

图2 两种不同网站访问实例RGB图像化纹理对比

Fig. 2 Comparison of RGB image textures from two different website visiting instances

2.2 模型选择

由于网站指纹特征是以RGB图像进行表示的，而卷积神经网络在图像分类问题上具有天然的优势，因此采用深层卷积神经网络作为Image-FP的分类模型。通过初步实验结果发现，相较于VGG网络或Inception网络，ResNet更能够从图像纹理中学习到特征模式进行分类。

ResNet由He等[13]提出，并在2015年的ImageNet大规模视觉识别挑战竞赛（ImageNet Large Scale Visual Recognition Challenge，ILSVRC）中获得第一名，同时模型的整体参数量相较于VGG网络更少，性能更加优异。

根据以往的经验，神经网络模型的结构越深，理论上模型的拟合性能会越好。然而在研究中发现，网络深度增加到某一程度后，其准确度会出现饱和的现象，甚至于开始下降，这就是退化问题（degradation problem）。

ResNet通过引入跨层连接技术，并加入了使用短路连接（shortcut connections）机制的残差块（residual block），很好地解决了退化问题。

图3展示了残差块的其中一种结构，可以看到残差块主要由两个权重层（weight layer）和两个激活层（relu）交替组成。假设残差块的输入为x，而结构中间的权重层待学习的部分为残差$F(x)$，则整个残差块的输出就由原始输入x及残差$F(x)$相加而成，即$F(x)+x$。这种短路设计使得就算权重层并没有学习到有用的相关特征（即$F(x)=0$），网络也能保持原来的性能而不会进一步降低。而实际情况是残差并不会真的为0，也就是说ResNet能够通过不断堆叠残差块来增加网络深度的同时，进而提升模型的拟合性能。

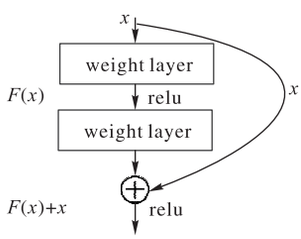


图3 残差块结构

Fig. 3 Structure of residual block

根据数据集规模选择了50层的ResNet网络，图4展示了网络的具体结构。RenNet50网络主要由两种结构块（conv_block和identity_block）交替堆叠组成。两种结构块左半部分相同，由三组的卷积层（Conv 3×3）、批规范化层（Batch Normalization）和激活层（Activation）组成。唯一不同之处在于右半部分的短路连接，conv_block在直接连接的基础上加入了卷积层（Conv 1×1）和批规范化层（Batch Normalization）。同时ResNet使用了全局平均池化（Average Pool）代替了全连接层，降低模型参数量的同时还能进一步提升网络的性能。

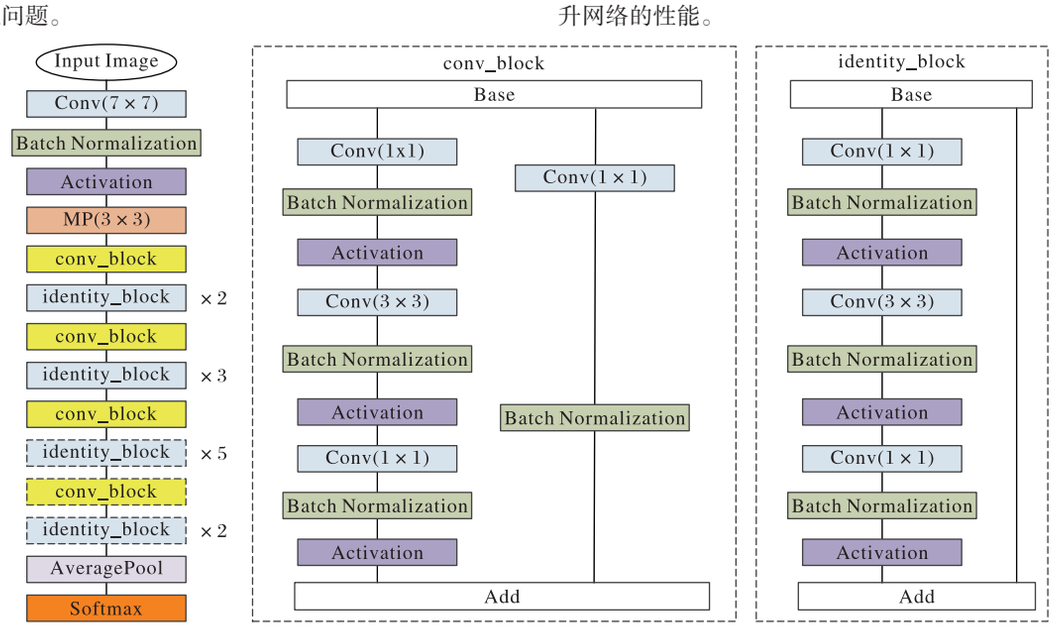


图4 Image-FP采用的ResNet50结构

Fig. 4 Structure of ResNet50 used in Image-FP

Image-FP保持ResNet50模型中默认的部分超参数，但是其中的权重参数会进行重新训练。选择Adamax作为模型的优化算法，并将其学习率设置为2×10^{-3}。Adamax是一种适应性矩估计的随机目标函数梯度优化算法，它通过对梯度的一阶矩估计（first moment estimation）和二阶矩估计（second moment estimation）进行综合考虑，从而计算出更新步长。Adamax具有计算高效和内存需求少等特点而适用于大量参数的神经网络模型中，同时也能有效解决梯度稀疏或梯度存

在高噪声的问题。最终模型选择如表1所示。

需要注意的是,ResNet接收的输入维度为224 × 224,因此会将原始1024 × 1024的RGB图像等比例缩放至该大小以满足神经网络的输入条件。

表1　Image-FP模型参数选择

Tab. 1　Model parameter selection of Image-FP

参数	选择范围	最终值
网络模型	VGG19,ResNet50,InceptionV3	ResNet50
优化器	SGD,RMSProp,Adam,Adamax	Adamax
学习率	0.001,0.002,…,0.01	0.002
学习率衰减	0.0,0.1,…,0.9	0.0
批大小	8,16,24,32	32

此外,Image-FP使用50层的卷积神经网络,远远大于其余两种基于深度学习方法的DF和AWF网站指纹技术,其卷积神经网络模型深度只有不到10层。因此实验结果也能表明神经网络模型的深度对于网站指纹攻击的性能影响。

3　实验与结果评估

3.1　实验环境

对于CUMUL、*k*-FP、AWF、DF和Image-FP这5种网站指纹技术,都使用Python(版本3.6)进行实现。其中:*k*-FP基于Python的机器学习库scikit-learn;深度学习模型(AWF、DF和Image-FP)基于Python的深度学习库来搭建,以Keras为前端、Tensorflow作为后端;而CUMUL则是基于LibSVM开源工具。

所有网站指纹模型都是在8核16线程的CPU上进行模型的训练和测试。对于深度学习模型,由于神经网络具有天然的并行计算的优势,因此使用显存8 GB的NVIDIA RTX 2070显卡进行GPU加速。

封闭世界和开放世界的数据集都会分成训练集和测试集两部分,其中:训练集包含80%的数据用于网站指纹分类器的训练,其余20%为测试集用来评估分类器的性能。同时对于每一种网站指纹技术,都会用10折交叉验证来计算平均的结果。

此外,对于AWF和DF这两种网站指纹技术,相较于文献[10-11]只使用±1的数据包方向作为神经网络模型的输入,发现使用包含数据包大小的时序序列±*PacketSize*对于网站指纹攻击的效果提升明显,因此后续实验将使用±*PacketSize*作为AWF和DF两种模型的输入。

3.2　封闭世界

封闭世界数据集由50个网站构成,其中每个网站有1 250条访问实例。按照训练集:测试集=8:2的原则,则对于每一个网站类别都有1 000条的数据用于分类器训练,其余250条数据作为测试集。

实验结果如表2所示,其中,Image-FP的网站指纹准确率在所有前沿的攻击方法中准确率是最高的,达到97.2%,相较于其余两种基于深度学习的方法DF和AWF,其准确率分别提高了0.4和1.3个百分点。实验结果表明,本文提出的基于图像纹理的网站指纹技术确实能够从原始的流量数据中学习到更多相关的抽象特征。同时,也可以发现,基于深度学习模型的网站指纹技术明显优于使用传统机器学习的方法(CUMUL和*k*-FP),在准确率上平均高出4个百分点,一定程度上表明卷积神经网络的自适应特征学习在网站指纹攻击上具备一定优势。

表2　网站指纹技术在封闭世界场景下的准确率　单位:%

Tab. 2　Accuracies of website fingerprinting techniques in closed-world scenario　unit:%

网站指纹技术	准确率	网站指纹技术	准确率
Image-FP	**97.2**	CUMUL	93.3
DF	96.8	*k*-FP	91.7
AWF	95.9		

进一步对Image-FP模型的收敛情况进行评估,也就是说:网络模型需要训练多久才能够从数据中学习到相关的特征模式。通常神经网络随着训练轮次的增加,模型的性能也会越来越好,但是也有可能会使模型陷入过拟合的情况,因此训练轮次对于模型的性能也有至关重要的影响。图5给出了Image-FP随着训练轮次增加在训练集和测试集上的性能。

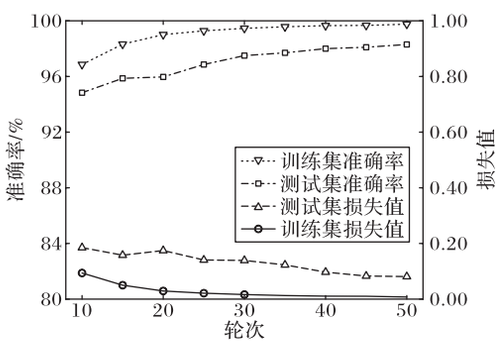


图5　Image-FP在训练集和测试集上的性能

Fig. 5　Performance of Image-FP on training set and testing set

由图5可以看到,在训练10轮左右,Image-FP已经可以达到接近95%的准确率;随着训练轮次的不断增加,Image-FP在40~50轮可以在测试集上取得最好的效果;而当训练轮次接近50轮时,Image-FP在训练集上的损失值已经接近于0,不过在测试集上的准确率没有进一步提升,表明模型已经达到饱和的状态。

图6给出了训练集大小对于不同网站指纹技术性能的影响。将每个网站的训练实例划分为5种情况,分别为200、400、600、800和1 000。由图6可以发现,当每个网站的训练实例只有200个时,Image-FP和DF已经能够取得90%以上的分类准确率,而AWF的准确率只有最低的73%。对于CUMUL、AWF和*k*-FP这3种网站指纹技术,它们分别需要400、600和700条训练数据才能达到相同的效果。

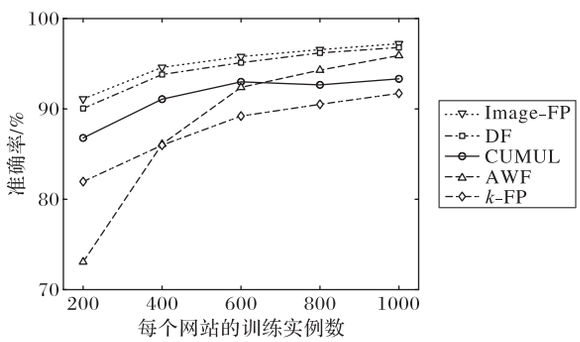


图6　训练集大小对网站指纹性能的影响(封闭世界)

Fig. 6　Impact of training dataset size on website fingerprinting performance (closed-world)

随着网站训练实例数量的增加,所有网站指纹技术的性

能也会不断提高，而Image-FP是当中表现最为优异的，并且其准确率也始终高于最先进的网站指纹技术DF。以上实验结果表明，Image-FP只需要少量的训练实例即可达到相当不错的效果，一定程度上降低了指纹攻击对数据规模的要求。

攻击者在实施网站指纹攻击前需要收集大量的训练数据，因此所耗费的计算资源和时间成本都需要考虑在内。文献[4,10,14]研究均表明，在网站指纹分类模型训练完成后的10~14 d，由于网站内容的变化，其攻击准确率会出现明显的下降，因此攻击者需要在数据训练集的规模以及更新频率上进行权衡。而只需少量的数据集就能表现出较好性能的Image-FP无疑是最好的选择。

从模型的训练时间上来说，DF、AWF和*k*-FP在较大规模的数据集上的训练速度都相对较快，Image-FP次之，而CUMUL的训练时间是所有网站指纹技术中最长的。进行50轮次训练的DF和AWF的平均训练时间在30~60 min，而Image-FP由于使用更大规模的输入数据和更深的神经网络，因此训练时间更长，需要大约5 h。另一方面，由于CUMUL需要对SVM分类器的超参数进行网格搜索来获得最优效果，因此往往需要花费数天的时间进行训练。此外，基于深度学习算法的3种网站指纹技术Image-FP、DF和AWF还具有分类预测较快的优点。

3.3 开放世界

相较于封闭世界，开放世界更加贴近于真实场景。在开放世界中，用户不再被限定只能访问数量有限的网站集合，而攻击者的目的在于从非监控网站的背景噪声流量中识别出用户访问的监控网站。

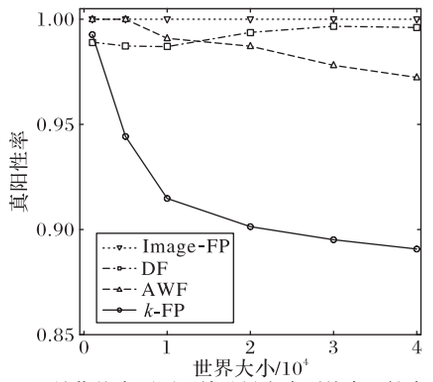
开放世界数据集包含62 500条监控网站实例和50 000条非监控网站实例。同样按照训练集∶测试集=8∶2的原则，因此训练集包含50 000条监控实例和40 000条非监控实例，而测试集包含12 500条监控实例和10 000条非监控实例。此外，将开放世界数据集中的监控和非监控网站数量统称为开放世界大小。

在开放世界中，网站指纹攻击是一个二元分类问题。因此，使用真阳性率(TPR)和假阳性率(FPR)来评估网站指纹攻击的效果。其中：TPR表示在所有实际上为监控网站实例的样本中，被网站指纹分类模型划分正确的比例；而FPR表示在所有实际上为非监控网站实例的样本中，被网站指纹分类模型错误划分为监控类别的比例。

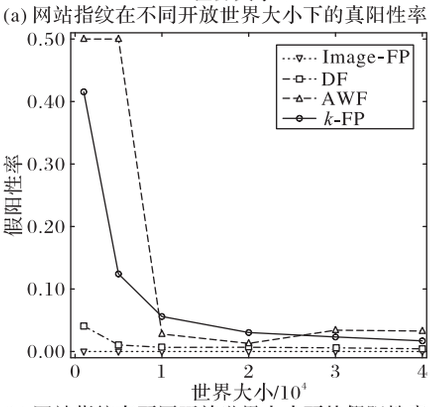
此外，由于CUMUL网站指纹技术无法适用于大规模数据的场景，因此在开放世界的实验中，本文只对其余4种网站指纹技术进行测试和评估。

图7给出了网站指纹技术在不同开放世界大小下的性能评估情况，其中Image-FP在所有网站指纹技术中表现最为出色。无论是在哪种世界大小下，Image-FP都能够以100%的准确率识别出监控网站的访问实例，并始终保持TPR为1和FPR为0。换句话说，Image-FP可以识别出全部的监控网站实例，同时也没有任何误判的情况。

相较于Image-FP，其余网站指纹技术的分类准确率都有不同程度的下降。随着世界大小的增加，DF的TPR不断提高，同时FPR也在同步下降，其性能相对接近于Image-FP。而AWF和*k*-FP两种网站指纹技术，尽管FPR随着非监控网站数量的增加而有大幅下降，但是它们的TPR却也在不断降低。*k*-FP在世界大小为40 000时甚至低于了0.9，表示该网站指纹技术并不适用于开放世界场景。



(a) 网站指纹在不同开放世界大小下的真阳性率



(b) 网站指纹在不同开放世界大小下的假阳性率

图7 开放世界大小对网站指纹性能的影响(开放世界)

Fig. 7 Impact of open-world size on website fingerprinting performance (Open World)

开放世界的实验结果表明，攻击者可以通过Image-FP网站指纹技术轻易地识别出目标用户访问的网站是否在自己的监控范围内，也一定程度上展示了Tor网络的匿名性在网站指纹攻击下的脆弱性。

4 结语

本文提出了一种全新的基于图像纹理和深度卷积神经网络的网站指纹技术Image-FP。Image-FP通过将Tor匿名通信流量按照二进制文件形式映射成RGB彩色图，再使用50层的深度卷积神经网络ResNet进行网站指纹特征的提取和分类。在封闭世界的场景下，Image-FP能够在最先进的网站指纹技术中取得最高的97.2%攻击准确率；而在更贴近真实环境的开放世界场景中，Image-FP能够以100%准确率识别出监控网站，是其他网站指纹技术所达不到的。实验结果表明，匿名流量图像化技术能够更广泛地保留网站指纹的相关特征，在网站指纹的刻画上更具有区分性，同时对Tor网络的匿名性带来了极大的挑战。

尽管本文提出的网站指纹技术能够在实验环境下取得较高的准确率，然而如文献[4,14]所述，实验环境与真实环境仍然存在较大的差距，因此在后续的研究中，还需对以下内容进行更深入的研究。

1)数据集的相似性。

在本文的实验中，匿名流量是在同一时间进行采集的，同时采集过程中的环境变量(如Tor浏览器版本、操作系统版

本、网络环境等)都相同或相似,因此训练集和测试集之间存在着较大的相似性,在一定程度上增强了网站指纹分类模型的性能,给予攻击者较大的优势条件。在更加真实的环境下,去掉上述前提条件对于Image-FP指纹技术的影响值得进一步测试和评估。

2)匿名流量分割。

本文在匿名流量数据的采集中,为了保证数据的纯净性,对用户加上了每次访问只能打开一个网站页面的前提条件。而在真实环境下,用户可以同时打开多个网页,而且网络中也会存在更多的背景噪声流量。因此在实际的攻击场景中,攻击者需要能够快速准确地过滤掉噪声流量,同时能够分割链路中多条不同的数据流。因此,匿名流量的分割技术也是亟待研究的重要课题之一。

3)网站整体的指纹刻画。

本文对于"网站"的定义与以往大多数的研究保持一致,即"网站"表示的是网站的主页,而不包括网站的其余子页面或是网站主页上的超链接。目前对于使用网站上的整体内容来对该网站进行完整的指纹刻画这一方面的研究较少,因此如何使用Image-FP的匿名流量图像化技术来对网站整体的指纹进行描述是本文后续的研究方向。

4)网站指纹防御技术。

尽管目前还并没有具体的攻击实例表明网站指纹攻击技术会对实际的Tor网络的匿名性造成破坏,研究者们也已经提出了不少的网站指纹防御技术[15-18]。这些防御技术通常采用对抗机器学习的方式来进行网站指纹攻击的缓解。在网络层面,发送虚假的数据包或延迟发送数据包都能有效加入噪声从而模糊流量当中的特征。发送虚假的数据包会增加Tor网络的链路负载,而延迟发送数据包会增加用户的等待时间从而影响用户体验。因此,如何设计网站指纹防御技术并能在安全和性能之间达到平衡,也是值得研究的方向。同时,Image-FP网站指纹技术在目前现有的防御策略上是否还能保持较高的分类准确率也需要进一步评估。

参考文献 (References)

[1] 何高峰,杨明,罗军舟,等. Tor匿名通信流量在线识别方法[J]. 软件学报,2013,24(3):540-556. (HE G F, YANG M, LUO J Z, et al. Online identification of Tor anonymous communication traffic [J]. Journal of Software, 2013, 24(3): 540-556.)

[2] 何永忠,李响,陈美玲,等. 基于云流量混淆的Tor匿名通信识别方法[J]. 工程科学与技术,2017,49(2):121-132. (HE Y Z, LI X, CHEN M L, et al. Identification of Tor anonymous communication with cloud traffic obfuscation [J]. Advanced Engineering Sciences, 2017, 49(2): 121-132.)

[3] PERRY M. A critique of website traffic fingerprinting attacks [EB/OL]. [2019-03-22]. https://blog. torproject. org/critique-website-traffic-fingerprinting-attacks.

[4] JUAREZ M, AFROZ S, ACAR G, et al. A critical evaluation of website fingerprinting attacks [C]// Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2014: 263-274.

[5] HERRMANN D, WENDOLSKY R, FEDERRATH H. Website fingerprinting: attacking popular privacy enhancing technologies with the multinomial naïve-Bayes classifier [C]// Proceedings of the 2009 ACM Workshop on Cloud computing security. New York: ACM, 2009: 31-42.

[6] PANCHENKO A, NIESSEN L, ZINNEN A, et al. Website fingerprinting in onion routing based anonymization networks [C]// Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society. New York: ACM, 2011: 103-114.

[7] WANG T, CAI X, NITHYANAND R, et al. Effective attacks and provable defenses for website fingerprinting [C]// Proceedings of the 23rd USENIX Security Symposium. Berkeley: USENIX Association, 2014: 143-157.

[8] PANCHENKO A, LANZE F, PENNEKAMP J, et al. Website fingerprinting at internet scale [EB/OL]. [2019-03-22]. https://www. comsys. rwth-aachen. de/fileadmin/papers/2016/2016-panchenko-ndss-fingerprinting. pdf.

[9] HAYES J, DANEZIS G. *k*-fingerprinting: a robust scalable website fingerprinting technique [C]// Proceedings of the 25th USENIX Security Symposium. Berkeley: USENIX Association, 2016: 1187-1203.

[10] RIMMER V, PREUVENEERS D, JUAREZ M, et al. Automated website fingerprinting through deep learning [EB/OL]. [2019-03-22]. https://arxiv. org/pdf/1708. 06376. pdf.

[11] SIRINAM P, IMANI M, JUAREZ M, et al. Deep fingerprinting: undermining website fingerprinting defenses with deep learning [C]// Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2018: 1928-1943.

[12] NATARAJ L, KARTHIKEYAN S, JACOB G, et al. Malware images: visualization and automatic classification [C]// Proceedings of the 8th International Symposium on Visualization for Cyber Security. New York: ACM, 2011: Article No. 4.

[13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

[14] WANG T, GOLDBERG I. On realistically attacking tor with website fingerprinting [J]. Proceedings on Privacy Enhancing Technologies, 2016(4): 21-36.

[15] DYER K P, COULL S E, RISTENPART T, et al. Peek-a-boo, I still see you: why efficient traffic analysis countermeasures fail [C]// Proceedings of the 2012 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2012: 332-346.

[16] CAI X, NITHYANAND R, WANG T, et al. A systematic approach to developing and evaluating website fingerprinting defenses [C]// Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2014: 227-238.

[17] JUAREZ M, IMANI M, PERRY M, et al. Toward an efficient website fingerprinting defense [C]// Proceedings of the 2016 European Symposium on Research in Computer Security, LNCS 9878. Cham: Springer, 2016: 27-46.

[18] WANG T, GOLDBERG I. Walkie-talkie: an efficient defense against passive website fingerprinting attacks [C]// Proceedings of the 26th USENIX Security Symposium. Berkeley: USENIX Association, 2017: 1375-1390.

ZHANG Daowei, born in 1993, M. S. candidate. His research interests include network security, deep learning.

DUAN Haixin, born in 1972, Ph. D., professor. His research interests include network security, network measurement.