> 一篇关于流量识别的粗浅个人总结
>
> 何正杰

流量识别技术概述

摘要

流量识别旨在从网络流量中分类出特定类别的流量，可以是特定应用的流量，应用的行为，恶意流量等等。流量识别作为一个传统的网络技术，已经研究了二十多年，并在入侵检测系统等安全应用中广泛使用。传统的流量识别根据端口号，协议等特征进行识别，但近年来，出于对安全和隐私的考虑，加密流量激增，为流量识别带来了新挑战。在本文中，我们对现有的流量分类技术进行了介绍，不仅包括传统的技术，还包括最近的技术和趋势。除此之外，我们还对流量分类任务根据其应用场景和特点进行了进一步的划分，讨论了适用于不同任务的方法。最后，我们讨论了流量分类未来可能的发展方向和挑战。

背景

迅猛增长的加密流量正不断改变着威胁形势。随着越来越多的企业实现全数字化，大量的服务和应用都采用加密技术作为确保信息安全的首要方法。更具体地说，加密流量同比增长已超过 90%，对流量进行加密的网站数量已从 2015 年的 21% 上升到 2016 年的超过 40%。据 Gartner 预测，到2020年，80% 以上的网站流量都会被加密。

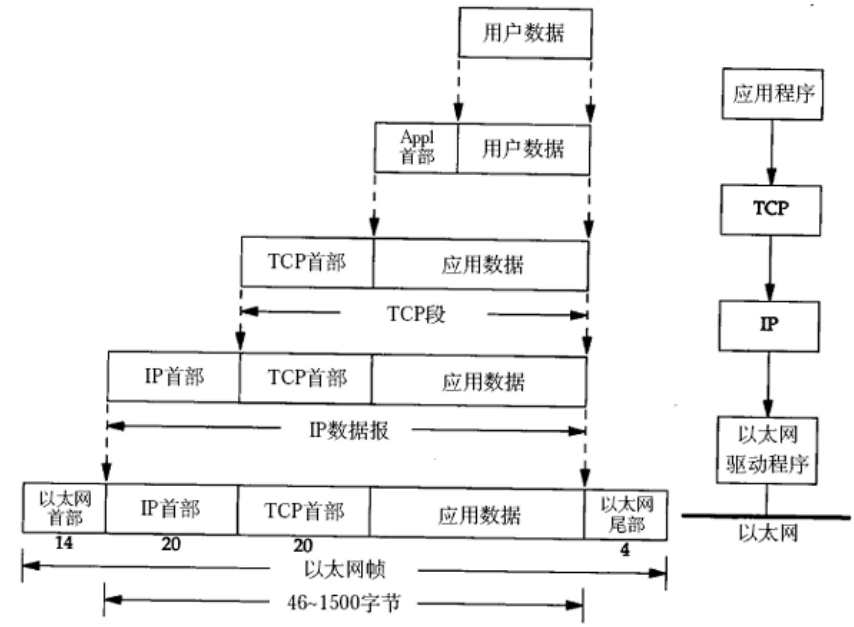
对于使用互联网通信并在线处理业务交易的企业而言，加密技术可以为其提供更强的隐私性和安全性。移动应用、云应用和 Web 应用依赖合理实施的加密机制，使用密钥和证书来确保安全性并建立信任。然而，企业并不是加密技术的唯一受益者。威胁发起者也在利用这项技术躲避检测，确保他们恶意活动能够得逞。

面对越来越多的恶意网络服务通过加密和隧道技术绕过防火墙和入侵检测系统，学术界和工业界急需对产生威胁的流量进行检测。

任务定义

流量识别旨在识别流量的意图，类别。在这里，流量指的是在网络中传输的数据包。

其源头是应用层的读写操作，经过传输层协议的变换（分片、协议状态机、加密等），流量序列产生一定变化。但是这种变化非常有限，因为流量的发生过程本质是确定性的，随机因素较小，因此对于特定环境中的特定应用（浏览器访问 google.com）各种流量特征体现出相当大的一致性和独特性，这就使“从流量特征识别应用”的监督学习问题成为可能。



根据流量意图的不同，流量在传输过程中会带有很多特征，一些工作利用单个数据包的特征对流量进行识别，一些工作则对整个网络数据包流进行处理，这会根据任务具体的目标不同而有所变化。

进行流量识别的第一步是明确识别的目标，对于不同的任务，由于数据的特点和流量的意图不同，所采用的方法也有所不同。下面我们从任务的目标角度，对常见的流量识别任务进行了划分。

应用流量分类

应用流量分类是流量识别中最常见的任务，该任务通常的做法是，对特定领域或者对常用的应用通过流量进行区分。通过对应用流量进行区分，一定程度上可以对用户隐私造成威胁，例如了解用户的医疗状况，性取向或宗教信仰。该任务在一些场景也有特定作用，比如俄罗斯，印度等国的流量审查。同时，由于目前加密流量检测难度的增大，网络黑产利用爬虫，刷流量，撸羊毛，扰乱正常网络秩序，为了解决这一问题，急需要更有效的手段对应用流量进行识别。

目前绝大部分的应用都通过加密流量进行数据传输，这使得过去利用IP五元组的方法失效，很多工作因此利用了统计特征和机器学习的方法进行流量识别，[Cliu2018]考虑流中包之间的传递关系的基础上，将多个应用的马尔科夫概率作为流的特征，在真实流量数据下仍能取得很好的效果。

对应用的流量分类不止局限于PC端，[Taylor2017]对Google Play商店中的110种最流行的应用进行了指纹识别，准确性达到了96%。[Sivanathan2019]使用诸如活动周期，端口号，信令模式和密码套件之类的统计特征对IOT设备进行了识别。一些工作还将应用流量识别拓展到了SDN上，[Cao2019]对SDN常见的10种软件在实验室环境中成功分类。这些工作证明了流量识别可以被广泛应用在各个场景下。

应用流量识别虽然在实验室往往可以取得90%以上的准确率，但是在现实中的效果会下降不少，这是因为实验室的环境中流量类别较少，噪声较少，并且现实世界中的数据分布是难以估计的，这与实验室数据集很可能不一致。即便如此，很多工作在现实环境中测试也能取得很好的效果。

网站指纹

网站指纹（website fingerprinting）的目标是根据流量对用户所访问的网页进行识别。最近有很多研究都表明，可以根据访问网站所产生的流量的数据包和大小检测特定的网站信息[Rimmer2018]。

在实验室的环境中，采用机器学习的网站指纹技术可以取得很好的效果，但是在真实环境中的效果往往会下降很多，这是因为很多研究的数据在实验室中是有限的，分类的类别也事先定义好，在面对真实世界大量的数据，且这个数据与实验室不一定是同分布的情况下，训练出来的分类器是否可用？这个问题尚有争论。

对于该任务而言，大多数工作采用的方法仍是提出特征再用机器学习训练分类器，例如[K. Liang2019]对于不同类型的特征使用特定类型的深度学习算法进行训练，将模型应用于网站服务分类中，一个网站服务可能包含多个网页，为用户提供该网站下的一种特定服务，该工作可以用于识别特定的网站服务。很多工作是在此流程的基础上，解决一些更细化的问题，比如[Zeng X2018]提出使用基于HTTP 1.1协议单 连接内串行请求特性刻画单个Web资源的局部特征模型，通过时间偏序关系的串联，提取前K个局部特征集，结合随机森 林算法，解决了云上单IP多网站难精细化区分的问题，实现指纹分类。[Rimmer2018]则利用匿名化网络tor的特性，采用长度序列的方向序列作为神经网络的输入，从而对访问的网页进行分类。

一些研究认为网站指纹攻击根据一定的改进在真是环境仍可行[Rimmer2018, Wang2016]，而一些研究[Panchenko2016]认为很多关于网站指纹的研究仅能识别一定的信息，无法很好的推广到开放世界。

应用行为分类

流量行为分类旨在识别流量的意图，该任务很多时候与应用分类同时进行，后者识别同类型的应用，前者在此基础上识别出应用的行为和步骤。行为分析可以暴露用户的隐私，分析用户的偏好，也可以在流量有一个更为全面的了解，对热点资源进行分析。从安全的角度而言，行为分类还可以识别出具有异常行为的流量，是一个值得研究的问题。

这里列举一些相关工作。

对于特定应用的流量识别会对用户隐私造成威胁，[Wang J2019]提取加密相机流量中的统计特征并使用机器学习方法对其进行分类，证实根据流量特征可识别用户行为进而构建日常生活行为规律。[Minghao2019]采用集成学习和决策树的方法对远程桌面的各个行为进行了流量识别，包括编辑文档，观看视频，安装软件等等，他们利用了远程桌面的一些常用协议作为特征。其他还有很多类似的工作比如[Xliu2019]对摄像头流量的用户行为进行了预测，这些工作往往是对特定功能的同类型APP的各种行为进行识别，并且可以利用其特有的协议作为特征。

混淆流量识别

很多根据统计特征进行流量识别的方法可以进流量进行很好的分类，暴露隐私，为了避免这一问题，可以采用流量混淆的方式对流量进行伪装，改变原本的特征[Wright2009]。流量混淆的方式包括且不限于：改变数据包大小，将数据包伪装成另外行为或应用的流量，伪装成不同的协议。

在[Dixon2016]的综述中提到上述流量伪装的方式在理论上可以被识别。对于填充的方式，虽然可以一定程度改变数据包的大小，但是会带来新的特征，并且诸如数据包时间戳等特征并不能进行伪装。在混淆流量识别方面有很多的工作，[xuemei2019]采用基于流上下文和主机流行为和DNS行为的方式对ShadowSocks匿名流量进行了识别。[Wang2015]采用信息熵的方式对采用填充的混淆流量进行了识别。对数据包进行变形同样也会有类似的问题，而模仿成其他协议的难点在于是否能进行完美的模仿。还有将流量伪装成没有任何特征的数据包，但并不能完美消除特征，有时没有特征也是一种特征。

同样的，目前的关于混淆流量的识别方式是否在真实环境中进行识别是有争议的。

恶意流量识别

越来越多的恶意网络服务通过加密和隧道技术绕过防火墙和入侵检测系统，这些恶意流量通过加密。隐藏了通讯内容，导致恶意流量很难被实时监控到，对互联网安全造成了严重的影响，及时发现恶意流量是学术界和工业界面临的一项巨大的挑战。

[Yu2019]认为用户欺诈的恶意机器人流量绝大多数来自数据中心，并提出了一种基于机器学习对恶意bot进行识别的准实时方法，在现实世界数据集中表现很好。

[Wang2018]通过将URL向量化对Android设备的恶意软件进行了分类。

[Shekhawat2019]也是通过机器学习的手段对恶意加密流量和良性流量进行了识别，他们采用了三种机器学习算法，分别是随机森林，Adaboost，SVM，值得一提的是，在这篇工作中，作者还对各种可利用的特征进行了整理。

类似利用统计特征和机器学习方法进行恶意流量识别的工作有很多，[Anderson2016]利用了多种统计特征如TLS握手元数据，链接到加密流的DNS上下文流以及5分钟内来自同一源IP地址的HTTP上下文流的HTTP头部字段采用有监督的机器学习对恶意流量进行了分类。[KLi2018]也是采用HTTP的统计特征来检测来自恶意软件的安全威胁。同样是利用HTTP统计特征，[Prassse2017]等人搜集了恶意应用与良性应用的数据流，利用基于域名的神经网络嵌入和处理网络流的长短期记忆网络进行恶意软件识别。

[Anderson2017]等人的工作现实了特征工程在通过机器学习识别恶意流量中起到了决定性作用，他们害比较了6中常见算法在面对真实环境下的性能，认为随机森林算法的效果在该任务中表现最为突出。

在检测恶意流浪时，采用机器学习的方法的工作比深度学习更多，这是因为恶意流量的数据是少量的，此时采用神经网络进行训练有过拟合的问题。

除了以上列出的任务之外，还有一些工作像[Roei2017]和[Ran2017]可以通过加密数据流识别出用户正在观看的视频内容。

[Martins2018]研究了流量拥塞问题，该工作训练了分类器对数据流分类，在SDN的环境下当第一个数据包到达时，可以预测流量特性采取合适的分流措施缓解拥塞问题。

流量识别的应用非常广泛，上面这些工作证明了虽然数据包进行了加密，我们的隐私并不是绝对安全的，流量的意图仍然有可能被识别。

常用方法总结

目前常用的流量识别方法就是采用特征工程的方式，对原始数据进行一定的预处理和清晰，根据任务目标抽取各种特征，采用专门设计的算法进行识别。过去比较常用的方式是采用字段匹配和正则表达式，在加密流量越来越普遍的现在，越来越多的工作利用统计特征，采用机器学习和深度学习的方式训练分类器进行识别，还有一些工作采用了半监督和无监督的方式。

在实验室环境中，为了从强调方法的效果，往往是采用单一的算法进行识别，在现实的应用场景中，工业界中网络流量分析（NTA）大多结合使用机器学习、高级分析和基于规则来检测企业网络上的可疑活动。

下面将对流量识别中的常用方法和特征进行介绍。

明文特征+规则匹配

传统报文检测会分析 IP 包的源地址、目的地址、源端口、目的端口以及协议类型。这种方式可以检查端口是否被正常使用，比如443端口是否传输的是明文流量。

DPI（Deep Packet Inspection）深度包检测技术是在传统报文检测之上增加了对应用层数据的协议识别，通过解析应用层数据来确定流量对应的业务信息。

传统报文检测和深度包检测都可以对流量进行一定程度的识别，但很多时候，对于加密流量和混淆后的流量难以检测。

有很多工作都是这种基于规则的思想，比如基于证书的检测，基于协议首部字段的检测，基于通讯模式的检测，对于明文数据是一种极为有效的手段，但人工分析特征需要专家得知识，并且很容易被拼接，改造后的 流量规避，对加密流量这样的手段起到的作用也是有限的。

近几年有不少工作都采用了这种类似规则匹配的方式，[Shbair2016]通过对比SNI和IP对应的域名信息来增强防火墙的安全性。[Husák2016]建立了SSL / TLS ciphersuite list和HTTP User-Agent的字典，并将User-Agent分配给观察到的SSL / TLS连接，用来识别通信的客户端。[Papadogiannaki2018]，从流量中提取行为特征，如包出现频率或包所在的位置，采用正则匹配固定模式，可以对facebook中的消息，语音和视频流量进行区别。

统计特征+机器学习

统计特征通常是对于整条数据流进行处理[Rezaei2018]，例如流长度，平均包长，流开始和结束时间等等，也可以跟多多个包之间的关系比如最小包间时延，包上下关联属性等。用统计特征不仅可以对明文传输的流量进行处理，还可以针对加密流量，基于统计特征的机器学习分类器已经成为识别加密流量的主要手段，但统计特征并非是完美的，比如混淆流量和恶意流量可以通过字段填充或伪装改变一定的统计特征。并且很多工作用到的统计特征需要对整条流进行处理，也就意味着很难做到实时的检测，如果仅能用于离线分类，应用场景会很有限。

除此之外，统计特征还需要人工设计和构建，依赖专业的知识和经验，这一点明文特征也要相同的缺陷，特征设计的好坏会影响最终识别的效果，如何构造的统计特征是这个问题的难点。

很多工作都是采用的统计特征，像[K. Li2018]采用HTTP请求的统计特征检测来自恶意软件的安全威胁，[Anderson2017]也是采用机器学习对统计特征训练了分类器，他们同时还考虑了带噪声的标签。

分类器的选择也会对识别的结果有着很大影响，比较常用的方法有决策树，随机森林，朴素贝叶斯，Adaboost，深度森林等等，这些算法有着各自的特点和优势，在[Namdev2015]这篇综述中，对机器学习的方法在流量识别中的应用进行了介绍。

大量数据+深度学习

流量进行加密处理后，调查基于规则匹配的方法已经逐渐失效，而许多工作中采用机器学习的方法训练分类器，需要人工提取特征，而这需要相关领域的专业知识，并且不同协议和应用构造的特征可能是不同的，在此基础上，一些工作将nlp中发展迅速的深度学习方法应用在了流量分类上，采用表示学习的思想可以避免繁琐的规则构造，从加密的原始信息中自动提取关键信息并生成有区分性的加密流量指纹。

对于采用深度学习方式进行流量识别的过程，[Rezaei2018]的综述中已经进行了较为详细的论述，简而言之，就是

在这方面也有很多经典的工作，[Lotfollahi2017]: 将payload从tcp/udp层开始对其填充，保证协议头部分20字节和8字节对齐，对于负载部分，采用前1480个字节，对于不足的负载进行0填充，保证维度一致性，使用ANN和SAE进行特征提取生成指纹。[Rimmer2018]:利用匿名化网络tor的特性，采用长度序列的方向序列作为深度学习网络SAE、CNN、LSTM的输入，从而分类访问的网页。[Liu2019]提出了一种基于表示学习的流序列网络，以双向GRU为基本单元，编码解码为整体结构，同时采用重构机制增强

加密流量指纹的表现能力，自动化的从原始流量信息中学习有效特征，提升了分类的精度。

尽管基于深度学习的识别方法在工作上显示出了很好的效果，但是这种方式往往是有监督的方式，需要大量的标注数据，对于数据的搜集而言是一个难点。

数据集介绍

很多工作都是采用自己收集的数据集，比如应用识别和行为识别往往都是针对特定的应用或者特定领域使用，所以没有公开的数据集资料，还有一些工作是利用一些软件生成的数据，由于收集软件的要求，不适合公开。由于真实网络环境下的流量是变化的，一些数据集比如KDDCUP99和NSL-KDD[Tavallaee2009]等已经和当前的真实网络环境差异较大。总体来说，在网络流量数据集方面，很难找到一个适用于多个领域的优秀数据集，但仍有一些工作将自己的数据集进行公开并且被广泛的使用，下面介绍加拿大网络安全研究所CIC的几个经典数据集，数据集可以用CIC提供的CICFlowMeter-V3对特征进行提取。。

[Gerard2016]加拿大网络安全研究所公布的VPN-nonVPN dataset (ISCXVPN2016)，数据集中包含7大类带标签的网络服务，数据是通过Wireshark和tcpdump捕获的，并且有VPN和nonVPN流量。

[Iman2018]通信安全机构(CSE)和加拿大网络安全研究所(CIC)之间的合作项目提供了CSE-CIC-IDS2018 on AWS数据集，该数据集包括七个不同的攻击场景：暴力攻击，Heartbleed，僵尸网络，DoS，DDoS，Web攻击以及内部网络的渗透，主要内容是每台计算机的网络流量和系统日志。

[Laya]安卓恶意流量数据集，主要包括5000多个软件的流量，其中有426个恶意软件，并且将这些恶意软件的流量分为了4类，广告软件，勒索软件，恐吓软件，短信恶意软件，除了网络流量外，还包括了各种日志文件和电池状态等等。

挑战和发展

流量识别是一个加密者与检测者对抗的过程，加密方采用各种加密手段和混淆手段消除特征，检测方也需要更前沿的方式，识别更有效地特征，构建新的模型进行识别。从趋势而言，更强的加密是未来的趋势，例如Tor的流量混淆插件OBFS，从obfs进化到obfs4，更强的加密使得特征难以被发现，给流量识别带来了更难的挑战。

在真实环境下实时检测

目前的很多工作都是在实验室的环境中测试，尤其是恶意流量识别和网站指纹等任务在真实环境中准确率和召回率都会下降很多。并且流量识别技术在落地部署时，还需要考虑实时性，这也就意味着基于数据流的方法都会失效，仅仅用几个数据包进行识别也会导致较高的假阳性。

比较典型的实时流量识别工作有，[Shim2017]提出了一种使用有效负载大小序列签名的应用分类方法，可以利用数据流前N个数据包的顺序，方向，有效负载大小生成负载大小序列签名进行应用分类。[Liu2017]利用时间切分的方式，将时间序列切分成小片段，利用碎片比对的方式实时分类。

[Rezaei2019]指出一些工作虽然用前几个数据包就能完成分类，但是这样会对造成巨大的负担，在一些应用场景中，只需要关注较长的流量。有些研究将整个流划分为几个部分，然后对每个部分进行分类，以检测不同的用户行为[Taylor2017]。

[Rezaei2018]的工作采用一种半监督方法，从而消除了对大型标记数据集的需求，每个类别仅用了20个样本，精度就可与大量标注数据的方法相比，并且通过采样的方法证明了从流的任意部分采样数据包足以进行分类。

更强加密协议的出现

协议的改进会使原本的一些特征和方法失效。在HTTP/1.1
中，只有前一个响应收到后才能发送下一个请求。为提高效率，HTTP/2提出复用和并发。浏览器不需等待接收上一个响应就可以启动下一个请求。服务端可以根据接收到的请求，发送任意一个响应数据包。这使同方向上连续非零负载数据包等burst特征失效。

针对QUIC的识别已经有一些工作做了尝试[Rezaei2018]并取得了不错的效果。

2018年8月，IETF正式发布TLS1.3协议的最终版本 (RFC 8446)，在安全性、性能和隐私等方面有重大改进，同时大大提升了HTTPS连接的速度。TLS1.3在TLS1.2的基础上做了改进，通过加密更多的握手过程(如证书交换) 来保护其免受窃听者的侵害，从而为数据交换提供更强的隐私性。ESNI（加密SNI）是TLS1.3协议的扩展，它可以阻止ISP、WiFi和其他监控者拦截TLS扩展模块中的SNI，防止用户访问网站的浏览记录被泄露，ESNI让使用HTTPS的互联网用户更难被跟踪。

除了新出现的通用协议外，一些恶意流量采用自定义的协议继续加密，例如APT攻击活动组织Wild Neutron使用的恶意软件与C&C服务器的通信都是使用自定义的协议进行加密，对于这类情况，特征很难提取到，并且数据收集较为困难，仍是一个值得研究的方向。

参考

本文主要参考了下面几篇综述或调研：

[klzgrad的流量分析survey][https://gist.github.com/klzgrad/73a6365c895e02d0e5d879cad1ffb691]

熊刚老师ISC2019的报告

Deep Learning for Encrypted Traffic Classification: An Overview

> 引用
>
> X. Zeng *et al*., "Flow Context and Host Behavior Based Shadowsocks's Traffic Identification," in *IEEE Access*, vol. 7, pp. 41017-41032, 2019, doi: 10.1109/ACCESS.2019.2907149.
>
> Dixon L, Ristenpart T, Shrimpton T. Network traffic obfuscation and automated internet censorship[J]. IEEE Security & Privacy, 2016, 14(6): 43-53.
>
> Rimmer V, Preuveneers D, Juarez M, et al. Automated feature extraction for website fingerprinting through deep learning[J]. arXiv preprint arXiv, 2017, 1708.
>
> Rezaei S, Liu X. How to achieve high classification accuracy with just a few labels: A semi-supervised approach using sampled packets[J]. arXiv preprint arXiv:1812.09761, 2018.
>
> Wang J, Cao Z, Kang C, et al. User Behavior Classification in Encrypted Cloud Camera Traffic[C]//2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019: 1-6.
>
> K. Liang, G. Gou, C. Kang, C. Liu, M. Yang and Y, Guo. "A Multi-view Deep Learning Model For Encrypted Website Service Classification," 2019 IEEE Global Communications Conference, 2019.

K. Li, R. Chen, L. Gu, C. Liu and J. Yin, "A Method Based on Statistical Characteristics for Detection Malware Requests in Network Traffic," 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, 2018, pp. 527-532, doi: 10.1109/DSC.2018.00084.

Anderson B, McGrew D. Machine learning for encrypted malware traffic classification: accounting for noisy labels and non-stationarity[C]//Proceedings of the 23rd ACM SIGKDD International Conference on knowledge discovery and data mining. 2017: 1723-1732.

Mohammad Lotfollahi, Ramin Shirali Hossein Zade, Mahdi Jafari Siavoshani, Mohammdsadegh Saberian. Deep Packet: A Novel Approach For Encrypted Traffic Classification Using Deep Learning. CoRR abs/1709.02656 (2017)

Ensafi R, Fifield D, Winter P, et al. Examining how the Great Firewall discovers hidden circumvention servers[C]//Proceedings of the 2015 Internet Measurement Conference. 2015: 445-458.

Namdev N, Agrawal S, Silkari S. Recent advancement in machine learning based internet traffic classification[J]. Procedia Computer Science, 2015, 60: 784-791.

Wright C V, Coull S E, Monrose F. Traffic Morphing: An Efficient Defense Against Statistical Traffic Analysis[C]//NDSS. 2009, 9.

C. Liu, Z. Cao, G. Xiong, G. Gou, S. Yiu and L. He, "MaMPF: Encrypted Traffic Classification Based on
Multi-Attribute Markov Probability Fingerprints," 2018 IEEE/ACM 26th International Symposium on
Quality of Service (IWQoS), Banff, AB, Canada, 2018.

Y.Guo, J.Shi, Z. Cao , ,C. Kang ,G. Xiong , and Z.Li ,"Machine Learning Based CloudBot Detection
Using Multi-layer Traffic Statistics" IEEE HPCC 2019 - IEEE High Performance Computing and Communications, Zhangjiajie, China, 2019.

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, 2017, 13(1): 63-78.

M. Jiang, G. Gou, J. Shi and G. Xiong, "I Know What You Are Doing With Remote Desktop," 2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC), London, United Kingdom, 2019, pp. 1-7, doi: 10.1109/IPCCC47392.2019.8958721.

X. Liu, J.Wang, Y. Yang and G. Xiong, " Inferring Behaviors via Encrypted Video Surveillance Traffic by
Machine Learning," IEEE HPCC 2019 - IEEE High Performance Computing and Communications,
Zhangjiajie, China, 2019.

Liu C, Cao Z, Xiong G, et al. Mampf: Encrypted traffic classification based on multi-attribute markov probability fingerprints[C]//2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). IEEE, 2018: 1-10.

A. Sivanathan et al., "Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics," in IEEE Transactions on Mobile Computing, vol. 18, no. 8, pp. 1745-1759, 1 Aug. 2019, doi: 10.1109/TMC.2018.2866249.

Cao J, Yang Z, Sun K, et al. Fingerprinting {SDN} Applications via Encrypted Control Traffic[C]//22nd International Symposium on Research in Attacks, Intrusions and Defenses ({RAID} 2019). 2019: 501-515.

R. Dubin, A. Dvir, O. Pele and O. Hadar, "I Know What You Saw Last Minute—Encrypted HTTP Adaptive Video Streaming Title Classification," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 12, pp. 3039-3049, Dec. 2017, doi: 10.1109/TIFS.2017.2730819.

Schuster R, Shmatikov V, Tromer E. Beauty and the burst: Remote identification of encrypted video streams[C]//26th {USENIX} Security Symposium ({USENIX} Security 17). 2017: 1357-1374.

Y. Guo, J. Shi, Z. Cao, C. Kang, G. Xiong and Z. Li, "Machine Learning Based CloudBot Detection Using Multi-Layer Traffic Statistics," 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Zhangjiajie, China, 2019, pp. 2428-2435, doi: 10.1109/HPCC/SmartCity/DSS.2019.00339.

Rezaei S, Liu X. Deep learning for encrypted traffic classification: An overview[J]. IEEE communications magazine, 2019, 57(5): 76-81.

Shim K S, Ham J H, Sija B D, et al. Application traffic classification using payload size sequence signature[J]. International Journal of Network Management, 2017, 27(5): e1981.

Liu J, Fu Y, Ming J, et al. Effective and real-time in-app activity analysis in encrypted internet traffic streams[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 335-344.

Wang S, Chen Z, Yan Q, et al. Deep and broad learning based detection of android malware via network traffic[C]//2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). IEEE, 2018: 1-6.

Shekhawat A S, Di Troia F, Stamp M. Feature analysis of encrypted malicious traffic[J]. Expert Systems with Applications, 2019, 125: 130-141.

K. Li, R. Chen, L. Gu, C. Liu and J. Yin, "A Method Based on Statistical Characteristics for Detection Malware Requests in Network Traffic," 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, 2018, pp. 527-532, doi: 10.1109/DSC.2018.00084.

Anderson B, McGrew D. Machine learning for encrypted malware traffic classification: accounting for noisy labels and non-stationarity[C]//Proceedings of the 23rd ACM SIGKDD International Conference on knowledge discovery and data mining. 2017: 1723-1732.

Prasse P, Machlica L, Pevný T, et al. Malware detection by analysing encrypted network traffic with neural networks[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2017: 73-88.

Anderson B, McGrew D. Identifying encrypted malware traffic with contextual flow data[C]//Proceedings of the 2016 ACM workshop on artificial intelligence and security. 2016: 35-46.

Martins R A S. Predicting Traffic Flow Size and Duration[D]. , 2018.

Shbair W M, Cholez T, François J, et al. Improving sni-based https security monitoring[C]//2016 IEEE 36th International Conference on Distributed Computing Systems Workshops (ICDCSW). IEEE, 2016: 72-77.

Gerard Drapper Gil, Arash Habibi Lashkari, Mohammad Mamun, Ali A. Ghorbani, "Characterization of Encrypted and VPN Traffic Using Time-Related Features", In Proceedings of the 2nd International Conference on Information Systems Security and Privacy(ICISSP 2016) , pages 407-414, Rome, Italy.

M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” *Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 2009.

Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018

Laya Taheri, Andi Fitriah Abdulkadir, Arash Habibi Lashkari; Extensible Android Malware Detection and Family Classification Using Network-Flows and API-Calls, The IEEE (53rd) International Carnahan Conference on Security Technology, India, 2019