(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 116233013 B
(45)授权公告日 2024.07.16

(21)申请号 202111470314.4

(22)申请日 2021.12.03

(65)同一申请的已公布的文献号
申请公布号 CN 116233013 A

(43)申请公布日 2023.06.06

(73)专利权人 南京理工大学
地址 210094 江苏省南京市玄武区孝陵卫200号

(72)发明人 刘伟伟 胡梁宏 曾盛 沈昊

(74)专利代理机构 南京理工大学专利中心 32203
专利代理师 朱炳斐

(51)Int.Cl.
H04L 47/2483(2022.01)
H04L 47/2441(2022.01)
H04L 41/14(2022.01)
H04L 43/045(2022.01)
H04L 43/10(2022.01)
H04L 9/40(2022.01)
H04L 12/46(2006.01)

(56)对比文件
CN 109951444 A,2019.06.28
CN 111224940 A,2020.06.02

审查员 朱文君

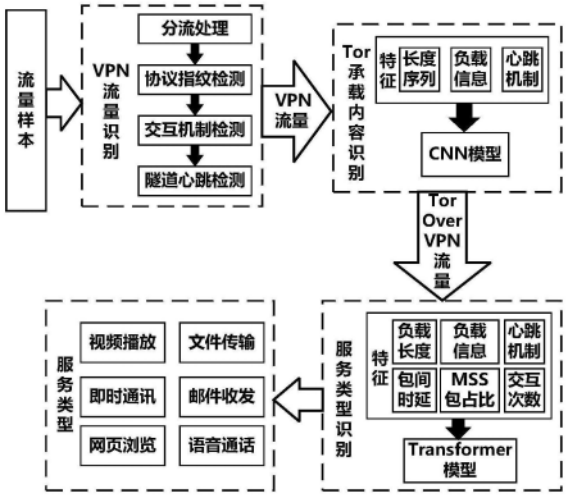
权利要求书2页 说明书4页 附图1页

(54)发明名称

针对Tor Over VPN匿名网络流量及其服务类型的识别方法

(57)摘要

本发明公开了一种针对Tor Over VPN匿名网络流量及其服务类型的识别方法。该方法包括:基于OpenVPN握手阶段的协议指纹、长度序列与心跳交互等特征筛选得到VPN流量;利用Tor网络建立阶段存在的握手长度序列以及心跳交互行为等机制,结合CNN模型识别出Tor Over VPN流量;最后,利用包间时延、包长度、包负载等特征结合Transformer深度学习模型识别Tor Over VPN流量的不同承载服务类型。本发明通过深入挖掘多维度时空特征、并结合当前主流深度学习模型,可以很好的对Tor Over VPN流量及其承载服务类型进行识别,对于加强Tor流量监管、维护网络安全具有重要意义。



CN 116233013 B

1.一种针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,所述方法包括以下步骤:

步骤1,基于五元组信息将输入流量样本分流处理为会话流量,并依据流量类型对其进行标记、分组、编号预处理操作;所述五元组为源地址、目的地址、源端口、目的端口、协议五元组;

步骤2,逐流提取序号为0至N1的数据包负载长度、OpenVPN头部协议字段、心跳数据包特征;

步骤3,逐流提取序号为N1至N2的数据包的负载长度、负载信息、轮询数据特征,并将这些特征转化为二维灰度图像;

步骤4,逐流提取序号为N2至N3数据包的长度、负载信息、包间时延、MSS包占比、交互次数,构成时空特征向量;

步骤5,利用步骤2提取的特征对流量进行匹配,识别OpenVPN隧道流量;

步骤6,基于所述二维灰度图像和CNN模型构建Tor Over VPN匿名网络流量识别模型;

步骤7,基于所述时空特征向量和Transformer模型构建服务类型识别模型;

步骤8,针对待检测的流量样本,执行步骤1至步骤4,之后根据步骤5识别OpenVPN隧道流量,利用步骤6和步骤7构建的模型分别识别Tor Over VPN匿名网络流量和服务类型。

2.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤1中对输入流量样本进行分流处理的同时,将MAC、IP、端口这些字节信息均置0。

3.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤1中进行编号之前,需去除负载长度为0的数据包。

4.根据权利要求1或3所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤1中所述编号遵循PSH标记位进行编号,满载包需与后续第一个非满载且PSH数据包组成一个分片。

5.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤2中序号为0至N1的数据包负载长度具体为:

逐流提取序号为0至N1的数据包负载长度,之后将其中固定不变的负载长度按序保留,作为0至N1的数据包负载长度特征,其余负载长度忽略,且将上行数据包长度标记为正数,下行数据包长度标记为负数。

6.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤3中将特征转化为二维灰度图像的过程中,若特征长度小于所设定的二维灰度图像的像素,进行填0补充,否则进行截断处理。

7.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤4中所述时空特征向量中的数据,依据数据包方向进行标记,上行数据包标记为正数,下行数据包标记为负数。

8.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤2中N1为30,N2为120,N3为320。

9.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤6所述基于所述二维灰度图像和CNN模型构建Tor Over VPN匿名网络流量

识别模型,具体包括:

对二维灰度图像添加Tor Over VPN匿名网络流量标签;

然后利用二维灰度图像对CNN模型进行训练,形成Tor Over VPN匿名网络流量识别模型。

10.根据权利要求1所述的针对Tor Over VPN匿名网络流量及其服务类型的识别方法,其特征在于,步骤7所述基于所述时空特征向量和Transformer模型构建服务类型识别模型,具体包括:

对时空特征向量添加服务类型标签;所述服务类型视频播放、文件传输、即时通讯、邮件收发、网页浏览、语音通话;

然后利用时空特征向量对Transformer模型进行训练,形成服务类型识别模型。

针对Tor Over VPN匿名网络流量及其服务类型的识别方法

技术领域

[0001] 本发明属于网络安全技术领域,特别是一种针对Tor Over VPN匿名网络流量及其服务类型的识别方法,通过提取Tor Over VPN流量的时空维度特征并结合CNN、Transformer等模型进行识别。

背景技术

[0002] Tor作为当前应用最广泛的匿名通信软件,通过三重加密转发技术实现代理传输用户数据,从而躲避网络安全机构审查。随着检测识别技术的提高,Tor原生模式以及混淆插件模式的使用体验日趋下降,用户逐渐转向Tor Over VPN模式,通过VPN前置代理的方式加密访问境外资源的同时保证匿名通信的目的。

[0003] Tor浏览器支持PC、Android、Linux等多个平台,可在官网免费下载安装使用,且各大论坛社区提供相关使用教程。此外,OpenVPN作为当前主流VPN,被各大社区推荐与Tor浏览器结合使用,用以更好的躲避ISP监管。

[0004] 目前,对Tor Over VPN模式流量的识别研究还较少,因此有必要针对Tor Over VPN匿名网络流量及其访问服务类型进行识别。

发明内容

[0005] 本发明的目的在于针对现有技术存在的问题,提供一种针对Tor Over VPN匿名网络流量及其服务类型的识别方法。

[0006] 实现本发明目的的技术解决方案为:一种针对Tor Over VPN匿名网络流量及其服务类型的识别方法,所述方法包括以下步骤:

[0007] 步骤1,基于五元组信息将输入流量样本分流处理为会话流量,并依据流量类型对其进行标记、分组、编号预处理操作;所述五元组为源地址、目的地址、源端口、目的端口、协议五元组;

[0008] 步骤2,逐流提取序号为0至N1的数据包负载长度、OpenVPN头部协议字段、心跳数据包特征;

[0009] 步骤3,逐流提取序号为N1至N2的数据包的负载长度、负载信息、轮询数据特征,并将这些特征转化为二维灰度图像;

[0010] 步骤4,逐流提取序号为N2至N3数据包的长度、负载信息、包间时延、MSS包占比、交互次数,构成时空特征向量;

[0011] 步骤5,利用步骤2提取的特征对流量进行匹配,识别OpenVPN隧道流量;

[0012] 步骤6,基于所述二维灰度图像和CNN模型构建Tor Over VPN匿名网络流量识别模型;

[0013] 步骤7,基于所述时空特征向量和Transformer模型构建服务类型识别模型;

[0014] 步骤8,针对待检测的流量样本,执行步骤1至步骤4,之后根据步骤5识别OpenVPN隧道流量,利用步骤6和步骤7构建的模型分别识别Tor Over VPN匿名网络流量和服务类

型。

[0015] 进一步地,步骤1中对输入流量样本进行分流处理的同时,将MAC、IP、端口这些字节信息均置0。

[0016] 进一步地,步骤1中进行编号之前,需去除负载长度为0的数据包。

[0017] 进一步地,步骤1中所述编号遵循PSH标记位进行编号,满载包需与后续第一个非满载且PSH数据包组成一个分片。

[0018] 进一步地,步骤2中序号为0至N1的数据包负载长度具体为:

[0019] 逐流提取序号为0至N1的数据包负载长度,之后将其中固定不变的负载长度按序保留,作为0至N1的数据包负载长度特征,其余负载长度忽略,且将上行数据包长度标记为正数,下行数据包长度标记为负数。

[0020] 进一步地,步骤3中将特征转化为二维灰度图像的过程中,若特征长度小于所设定的二维灰度图像的像素,进行填0补充,否则进行截断处理。

[0021] 进一步地,步骤4中所述时空特征向量中的数据,依据数据包方向进行标记,上行数据包标记为正数,下行数据包标记为负数。

[0022] 进一步地,步骤6所述基于所述二维灰度图像和CNN模型构建Tor Over VPN匿名网络流量识别模型,具体包括:

[0023] 对二维灰度图像添加Tor Over VPN匿名网络流量标签;

[0024] 然后利用二维灰度图像对CNN模型进行训练,形成Tor Over VPN匿名网络流量识别模型。

[0025] 进一步地,步骤7所述基于所述时空特征向量和Transformer模型构建服务类型识别模型,具体包括:

[0026] 对时空特征向量添加服务类型标签;所述服务类型视频播放、文件传输、即时通讯、邮件收发、网页浏览、语音通话;

[0027] 然后利用时空特征向量对Transformer模型进行训练,形成服务类型识别模型。

[0028] 本发明与现有技术相比,其显著优点为:

[0029] 1)通过挖掘VPN隧道建立阶段的长度序列、协议指纹与心跳机制等特征,基于规则匹配方式实现对OpenVPN隧道流量的检测,提高了对加密隧道流量检测的准确率。

[0030] 2)通过将负载长度、载荷信息、轮询数据等属性转化为灰度图的形式,并结合CNN模型实现对Tor Over VPN流量的识别,提高了对Tor Over VPN流量的检测能力。

[0031] 3)针对Tor Over VPN流量的承载服务类型识别问题,多维度深层次分析各类流量的时空特性分布差异,选取有效特征构建统一特征向量并结合Transformer时空序列模型对其进行精细化识别,提高了对Tor行为层面的服务类型识别精度。

[0032] 总体来说,本发明通过深入挖掘多维度时空特征、并结合当前主流深度学习模型,可以很好的对Tor Over VPN流量及其承载服务类型进行识别,对于加强Tor流量监管、维护网络安全具有重要意义。

[0033] 下面结合附图对本发明作进一步详细描述。

附图说明

[0034] 图1为本发明针对Tor Over VPN匿名网络流量识别方法的流程示意图。

[0035] 图2为OpenVPN协议格式图。

具体实施方式

[0036] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0037] 图1为本发明针对Tor Over VPN匿名网络流量的识别方法的流程示意图。该检测方法包括以下步骤:

[0038] 输入样本流量为各类协议流量,其中包括Tor Over VPN匿名网络流量;

[0039] VPN流量识别:首先基于五元组与协议进行分流处理,然后基于图2的OpenVPN协议格式、VPN握手建立交互机制与心跳保活等特征对输入流量样本进行识别,输出OpenVPN隧道流量;

[0040] Tor Over VPN流量识别:利用Tor网络建立过程中的长度序列、负载信息、轮询查询机制等特征结合CNN模型对VPN流量的承载内容进行识别,输出Tor Over VPN匿名网络流量;

[0041] 服务类型识别:在识别出Tor Over VPN流量的基础上,利用负载信息、心跳机制、包间时延、MSS满载包占比等时空维度特征并结合Transformer模型对使用Tor Over VPN访问不同目标过程中传输的应用数据类别进行识别,识别的类型包括视频播放、文件传输、即时通讯、邮件收发、网页浏览、语音通话等行为。

[0042] 具体地,在一个实施例中,提供了一种针对Tor Over VPN匿名网络流量及其服务类型的识别方法,所述方法包括以下步骤:

[0043] 步骤1,基于五元组信息将输入流量样本分流处理为会话流量,并依据流量类型对其进行标记、分组、编号预处理操作;所述五元组为源地址、目的地址、源端口、目的端口、协议五元组;

[0044] 这里,对输入流量样本进行分流处理的同时,将MAC、IP、端口这些字节信息均置0。

[0045] 这里,进行编号之前,需去除负载长度为0的数据包。

[0046] 这里优选地,所述编号遵循PSH标记位进行编号,满载包需与后续第一个非满载且PSH数据包组成一个分片。

[0047] 步骤2,逐流提取序号为0至30的数据包负载长度、OpenVPN头部协议字段、心跳数据包特征;

[0048] 这里,序号为0至30的数据包负载长度具体为:

[0049] 逐流提取序号为0至30的数据包负载长度,之后将其中固定不变的负载长度按序保留,作为0至30的数据包负载长度特征,其余负载长度忽略,且将上行数据包长度标记为正数,下行数据包长度标记为负数。

[0050] 步骤3,逐流提取序号为30至120的数据包的负载长度、负载信息、轮询数据特征,并将这些特征转化为二维灰度图像;

[0051] 这里,转化为二维灰度图像的过程中,若特征长度小于所设定的二维灰度图像的像素,进行填0补充,否则进行截断处理。

[0052] 步骤4,逐流提取序号为120至320数据包的长度、负载信息、包间时延、MSS包占比、

交互次数,构成时空特征向量;所述时空特征向量中的数据,依据数据包方向进行标记,上行数据包标记为正数,下行数据包标记为负数。

[0053] 步骤5,利用步骤2提取的特征对流量进行匹配,识别OpenVPN隧道流量;

[0054] 步骤6,基于所述二维灰度图像和CNN模型构建Tor Over VPN匿名网络流量识别模型,具体过程包括:

[0055] 对二维灰度图像添加Tor Over VPN匿名网络流量标签;

[0056] 然后利用二维灰度图像对CNN模型进行训练,形成Tor Over VPN匿名网络流量识别模型。

[0057] 步骤7,基于所述时空特征向量和Transformer模型构建服务类型识别模型,具体包括:

[0058] 对时空特征向量添加服务类型标签;所述服务类型视频播放、文件传输、即时通讯、邮件收发、网页浏览、语音通话;

[0059] 然后利用时空特征向量对Transformer模型进行训练,形成服务类型识别模型。

[0060] 步骤8,针对待检测的流量样本,执行步骤1至步骤4,之后根据步骤5识别OpenVPN隧道流量,利用步骤6和步骤7构建的模型分别识别Tor Over VPN匿名网络流量和服务类型。

[0061] 本发明通过深入挖掘多维度时空特征、并结合当前主流深度学习模型,可以很好的对Tor Over VPN流量及其承载服务类型进行识别,对于加强Tor流量监管、维护网络安全具有重要意义。

[0062] 以上显示和描述了本发明的基本原理、主要特征及优点。本行业的技术人员应该了解,本发明不受上述实施例的限制,上述实施例和说明书中描述的只是说明本发明的原理,在不脱离本发明精神和范围的前提下,本发明还会有各种变化和改进,这些变化和改进都落入要求保护的本发明范围内。本发明要求保护范围由所附的权利要求书及其等效物界定。

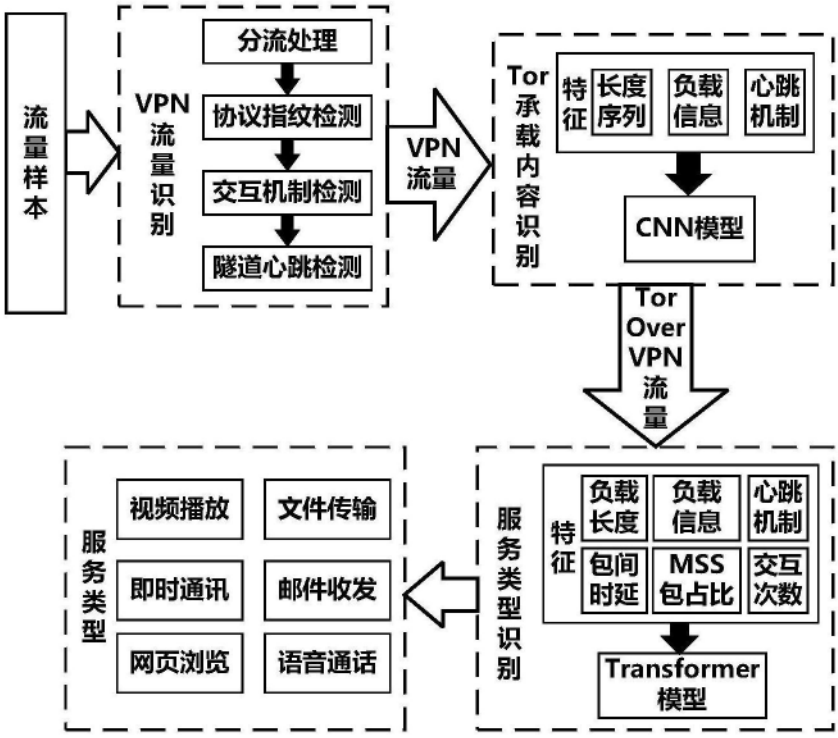

流量样本
VPN流量识别
分流处理
协议指纹检测
交互机制检测
隧道心跳检测
VPN流量
Tor承载内容识别
特征
长度序列
负载信息
心跳机制
CNN模型
Tor Over VPN流量
服务类型识别
特征
负载长度
负载信息
心跳机制
包间时延
MSS包占比
交互次数
Transformer模型
服务类型
视频播放
文件传输
即时通讯
邮件收发
网页浏览
语音通话

图1


0bit
8bit
16bit
24bit
32bit
Packet Length
Type
Session ID
Message Packet-ID Array Length
Message Packet-ID
Payload
注释
Packet Length：数据包负载长度
Type：数据包类型
Session ID：会话ID值
Message Packet-ID Array Length：Packet-ID数组长度
Message Packet-ID：Packet-ID值
Payload：加密负载数据

图2