(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108924090 A
(43)申请公布日 2018.11.30

(21)申请号 201810565176.X

(22)申请日 2018.06.04

(71)申请人 上海交通大学
地址 200240 上海市闵行区东川路800号

(72)发明人 邹福泰 朱宸 熊瑶庭 李林森 吴越 齐开悦 易平

(74)专利代理机构 上海旭诚知识产权代理有限公司 31220
代理人 郑立

(51)Int.Cl.
H04L 29/06(2006.01)
H04L 12/26(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

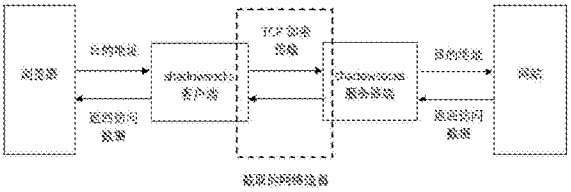
权利要求书1页 说明书3页 附图1页

(54)发明名称

一种基于卷积神经网络的shadowsocks流量检测方法

(57)摘要

本发明公开了一种基于卷积神经网络的shadowsocks流量检测方法，涉及计算机网络安全领域，包括以下步骤：通过抓包工具获取shadowsocks与普通流量；以TCP流为单位将流量拆分；提取TCP流的有效载荷，并拼接在一起，提取为十进制数，作为训练数据参数；对每个TCP流判断是否为shadowsocks流量，并进行数据标注，作为训练数据的结果；将标注好的训练数据作为卷积神经网络模型的训练输入，对模型进行训练，得出最终的计算模型。本发明将网络流的有效载荷转化为类似像素点的数据，将网络流量转化为图像，输入到CNN算法中。该方法省去了对流量进行特征提取的步骤，解决了无法找到shadowsocks流量决定性特征的问题。



CN 108924090 A

1.一种基于卷积神经网络的shadowsocks流量检测方法，其特征在于，包括以下步骤：

步骤1、通过抓包工具获取shadowsocks与普通流量；

步骤2、以TCP流为单位将流量拆分；

步骤3、提取TCP流的有效载荷，并拼接在一起，提取为十进制数，作为训练数据参数；

步骤4、对每个TCP流判断是否为shadowsocks流量，并进行数据标注，作为训练数据的结果；

步骤5、将标注好的训练数据作为卷积神经网络模型的训练输入，对模型进行训练，得出最终的计算模型。

2.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述步骤1中的shadowsocks流量和普通流量均大于1GB。

3.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述步骤1还包括以下步骤：

步骤1-1、shadowsocks将网络请求传输到本地服务器；

步骤1-2、经由加密后通过TCP连接与服务端进行通信，捕获到的shadowsocks流量是各个加密后的TCP流的集合。

4.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述步骤2中的流量拆分是指从双方TCP链接的发起到结束作为一个流，把2GB的流量拆分开来，便于进行数据的提取和标注。

5.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述步骤3还包括以下步骤：

步骤3-1、将获得每个流的实际传输内容，即有效载荷，取前1024位十六进制位；

步骤3-2、以两位16进制数为单位将流数据转化为512个0到255的十进制数，作为卷积神经网络的参数输入。

6.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述数据标注是用0代表非shadowsocks流量，用1代表shadowsocks流量。

7.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述shadowsocks的客户端与远程服务器之间利用TCP流进行传输。

8.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述shadowsocks流量采取不同的加密-解密方式。

9.如权利要求1所述的基于卷积神经网络的shadowsocks流量检测方法，其特征在于，所述最终的计算模型能够对实时网络流进行判定，预测是否为shadowsocks流量。

一种基于卷积神经网络的shadowsocks流量检测方法

技术领域

[0001] 本发明涉及计算机网络安全领域，尤其涉及一种基于卷积神经网络的shadowsocks流量检测方法。

背景技术

[0002] Shadowsocks是一种基于SOCKS5的加密代理工具。该工具在SOCK5协议基础上进行了加密和重构，用以隐匿客户端与代理服务器之间的传递内容，实现了高安全性和隐蔽性。目前该工具使用Python、C、C++、C#、Go语言等编程语言开发，分为客户端和服务端两个部分。该工具的运行流程如下：1、本地浏览器将请求通过SOCKS5协议交给本地的服务器的1080端口去代理。2、本地服务器运行local.py并监听1080端口，接受来自浏览器的请求。3、local.py接收到请求后将流量加密，通过TCP连接传输到shadowsocks远程服务器端。4、远程服务器解密请求后，访问浏览器将要请求的目的Ip。5、远程服务器将目的ip返回的内容加密后返回本地服务器。

[0003] 本地与远程服务器的加密基于服务器设置的用户名和密码进行对称加密，所以在传输过程中不需要进行密钥的交换。双方之间的通信是普通的TCP传输，与普通https流量在内容上没有任何差异，因此具备极高的隐蔽性。此外，由于此代理工具的隐蔽性，国内的上网用户常常通过购买国外服务器后搭建shadowsocks服务端，用于逃过国内的网络监管，从而非法访问境外网站。

[0004] 由于该工具产生的网络流量与普通加密流量没有本质上的区别，因此很难通过人工设计算法去进行特征识别。目前已有的识别shadowsocks的方式主要基于机器学习中的监督式或半监督式学习。例如利用随机森林算法对网络流的特征进行学习，从而试图找出shadowsocks流量与普通流量的特征差别。然而这些方式得到的识别准确率不高，容易将普通流量误认为shadowsocks流量。因此，这种方法很难应用到实际的网络监管中。

[0005] 机器学习中，卷积神经网络(Convolutional Neural Network,CNN)是一种基于前馈神经网络的算法，在近年由于其高效性引起了广泛关注。其在大型图像处理中往往有比较优秀的表现。本发明将网络流中的有效载荷(payload)转换为类似图像的输入，然后利用CNN算法训练模型。

[0006] 因此，本领域的技术人员致力于开发一种基于卷积神经网络的shadowsocks流量检测方法，从而解决常用的机器学习方法无法找到shadowsocks流量决定性特征的问题。

发明内容

[0007] 有鉴于现有技术的上述缺陷，本发明所要解决的技术问题是克服无法找到shadowsocks流量决定性特征的缺陷，目的在于提出一种基于卷积神经网络(CNN)的shadowsocks流量检测方法。采用shadowsocks与非shadowsocks的网络流，利用CNN算法训练模型，随后将该模型应用到实时的流量监测中。

[0008] 为实现上述目的，本发明提供了一种基于卷积神经网络的shadowsocks流量检测

方法，包括以下步骤：

[0009] 步骤1、通过抓包工具获取shadowsocks与普通流量；

[0010] 步骤2、以TCP流为单位将流量拆分；

[0011] 步骤3、提取TCP流的有效载荷，并拼接在一起，提取为十进制数，作为训练数据参数；

[0012] 步骤4、对每个TCP流判断是否为shadowsocks流量，并进行数据标注，作为训练数据的结果；

[0013] 步骤5、将标注好的训练数据作为卷积神经网络模型的训练输入，对模型进行训练，得出最终的计算模型。

[0014] 进一步地，所述步骤1中的shadowsocks流量和普通流量均大于1GB。

[0015] 进一步地，所述步骤1还包括以下步骤：

[0016] 步骤1-1、shadowsocks将网络请求传输到本地服务器；

[0017] 步骤1-2、经由加密后通过TCP连接与服务端进行通信，捕获到的shadowsocks流量是各个加密后的TCP流的集合。

[0018] 进一步地，所述步骤2中的流量拆分是指从双方TCP链接的发起到结束作为一个流，把2GB的流量拆分开来，便于进行数据的提取和标注。

[0019] 进一步地，所述步骤3还包括以下步骤：

[0020] 步骤3-1、将获得每个流的实际传输内容，即有效载荷，取前1024位十六进制位；

[0021] 步骤3-2、以两位16进制数为单位将流数据转化为512个0到255的十进制数，作为卷积神经网络的参数输入。

[0022] 进一步地，所述数据标注是用0代表非shadowsocks流量，用1代表shadowsocks流量。

[0023] 进一步地，所述shadowsocks的客户端与远程服务器之间利用TCP流进行传输。

[0024] 进一步地，所述shadowsocks流量采取不同的加密-解密方式。

[0025] 进一步地，所述最终的计算模型能够对实时网络流进行判定，预测是否为shadowsocks流量。

[0026] 本发明的基于CNN的shadowsocks流量检测方法，将网络流的有效载荷转化为类似像素点的数据，将网络流量转化为图像，输入到CNN算法中。该方法省去了对流量进行特征提取的步骤，解决了无法找到shadowsocks流量决定性特征的问题。

[0027] 以下将结合附图对本发明的构思、具体结构及产生的技术效果作进一步说明，以充分地了解本发明的目的、特征和效果。

附图说明

[0028] 图1是本发明的一个较佳实施例的捕获的shadowsocks流量的传输过程示意图；

[0029] 图2是本发明的一个较佳实施例的流数据处理过程示意图。

具体实施方式

[0030] 以下参考说明书附图介绍本发明的多个优选实施例，使其技术内容更加清楚和便于理解。本发明可以通过许多不同形式的实施例来得以体现，本发明的保护范围并非仅限

于文中提到的实施例。

[0031] 在附图中，结构相同的部件以相同数字标号表示，各处结构或功能相似的组件以相似数字标号表示。附图所示的每一组件的尺寸和厚度是任意示出的，本发明并没有限定每个组件的尺寸和厚度。为了使图示更清晰，附图中有些地方适当夸大了部件的厚度。

[0032] 本发明提供了一种基于CNN的shadowsocks流量检测方法。该方法由训练和预测两部分组成，具体包括以下步骤：

[0033] 1)抓取海量的shadowsocks与普通流量：由于shadowsocks的客户端与远程服务器之间利用TCP流进行传输，因此也需要抓取普通的TCP流量进行训练集的制作。另外，由于shadowsocks流量会采取不同的加密-解密方式，同样需要抓取各种类型的大量shadowsocks流量来确保shadowsocks流量的一般性。在此分别抓取大于1G的shadowsocks流量与普通TCP流量。

[0034] 如图1所示，是捕获的shadowsocks流量的传输过程，shadowsocks将网络请求传输到本地服务器，经由加密后通过TCP连接与服务端进行通信，捕获到的shadowsocks流量将是各个加密后的TCP流的集合。

[0035] 2)将流量拆分为流：从双方TCP链接的发起到结束作为一个流，把2G的流量拆分开来，便于进行数据的提取和标注。

[0036] 3)提取出每个流的有效载荷并拼接在一起。

[0037] 如图2所示，是对流数据进行处理的过程。首先将获得每个流的实际传输内容，即有效载荷(payload)，取前1024位十六进制位，然后以两位16进制数为单位将流数据转化为512个0到255的十进制数，作为卷积神经网络的参数输入。

[0038] 4)对每个流进行数据标注，0代表非shadowsocks流量，1代表shadowsocks流量。

[0039] 5)将结果输入到CNN算法中进行训练，得出最终的计算模型。该模型将用于shadowsocks流量的检测。

[0040] 训练完成后，将利用训练好的模型对实时网络流进行判定，预测是否为shadowsocks流量。

[0041] 以上详细描述了本发明的较佳具体实施例。应当理解，本领域的普通技术无需创造性劳动就可以根据本发明的构思作出诸多修改和变化。因此，凡本技术领域中技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案，皆应在由权利要求书所确定的保护范围内。

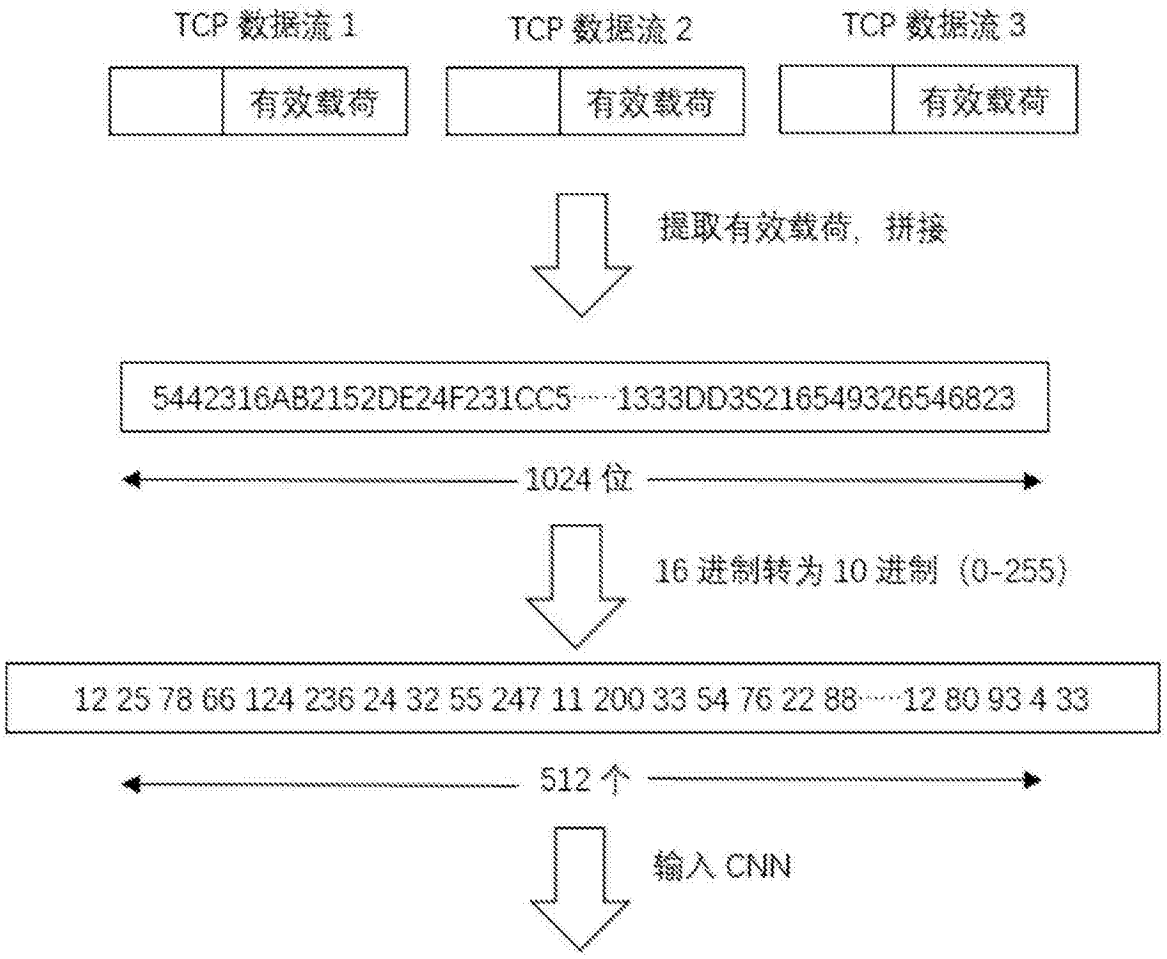

TCP 数据流 1
TCP 数据流 2
TCP 数据流 3
有效载荷
有效载荷
有效载荷
提取有效载荷，拼接
5442316AB2152DE24F231CC5……1333DD3S216549326546823
1024 位
16 进制转为 10 进制（0-255）
12 25 78 66 124 236 24 32 55 247 11 200 33 54 76 22 88……12 80 93 4 33
512 个
输入 CNN

图1

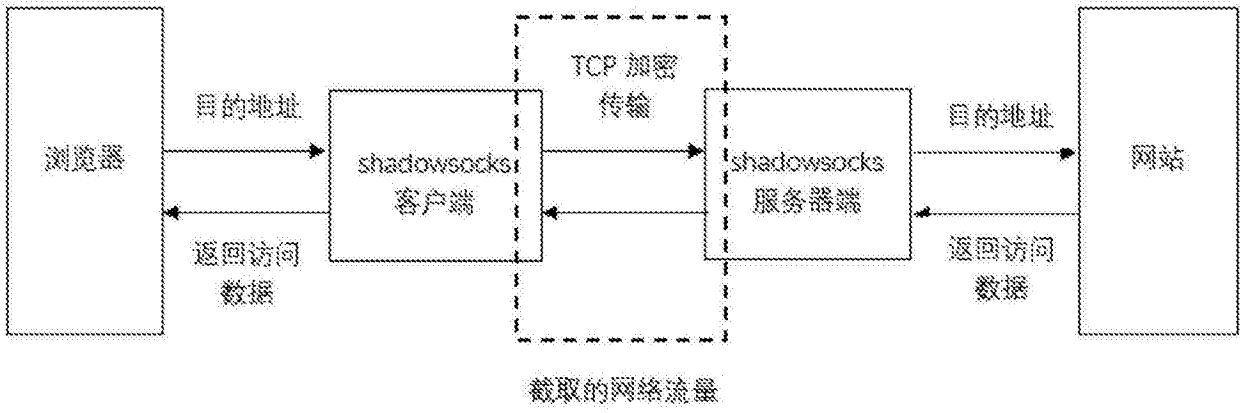

浏览器
目的地址
返回访问数据
shadowsocks 客户端
TCP 加密传输
shadowsocks 服务器端
目的地址
返回访问数据
网站
截取的网络流量

图2