(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115225575 B
(45)授权公告日 2023.11.24

(21)申请号 202210641983.1

(22)申请日 2022.06.08

(65)同一申请的已公布的文献号
申请公布号 CN 115225575 A

(43)申请公布日 2022.10.21

(73)专利权人 香港理工大学深圳研究院
地址 518057 广东省深圳市南山区粤海街道高新技术产业园南区粤兴一道18号香港理工大学产学研大楼205室

(72)发明人 王丹

(74)专利代理机构 深圳市君胜知识产权代理事务所(普通合伙) 44268
专利代理师 刘芙蓉 李可

(51)Int.Cl.
H04L 47/10(2022.01)
H04L 47/2441(2022.01)

(56)对比文件
CN 107301353 A,2017.10.27
WO 2022052476 A1,2022.03.17
WO 2021213123 A1,2021.10.28
CN 111970277 A,2020.11.20
CN 113537509 A,2021.10.22
CN 113361694 A,2021.09.07
US 11145020 B1,2021.10.12
Iman Akban.Encrypted web traffic classification using deep learning.UWSpace.2021,全文.
Dinh C. Nguyen .Federated Learning for Internet of Things: A Comprehensive Survey.IEEE.2021,全文.
Jasim, K.A Framework for Detection and Identification the Components of Arguments in Arabic Legal Texts.2019 first international conference of computer and applied science.2019,全文.
Jiahui Geng .DID-eFed: Facilitating Federated Learning as a Service with Decentralized Identities.ACM.2021,全文.

审查员 曹彦

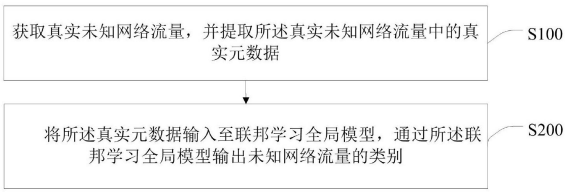
权利要求书2页 说明书8页 附图3页

(54)发明名称

一种基于元数据辅助和联邦学习的未知网络流量分类方法

(57)摘要

本发明公开了一种基于元数据辅助和联邦学习的未知网络流量分类方法,所述方法包括:获取真实未知网络流量,并提取所述真实未知网络流量中的真实元数据;其中,所述真实元数据为网络流量数据的部分字节数据;将所述真实元数据输入至联邦学习全局模型,通过所述联邦学习全局模型输出未知网络流量的类别;其中,所述联邦学习全局模型基于若干已训练的二分类器得到。本发明实施例基于元数据和二分类器,根据联邦学习全局模型来对未知网络流量进行分类识别,使得本发明能够一个具有未知网络流量的客户端在保护数据隐私和安全的情况下能从其他客户端学习到未知流量的分类方法。



CN 115225575 B

1.一种基于元数据辅助和联邦学习的未知网络流量分类方法,其特征在于,所述方法包括:

获取真实未知网络流量,并提取所述真实未知网络流量中的真实元数据;其中,所述真实元数据为网络流量数据的部分字节数据;

将所述真实元数据输入至联邦学习全局模型,通过所述联邦学习全局模型输出未知网络流量的类别;其中,所述联邦学习全局模型基于若干已训练的二分类器得到;

将所述真实未知网络流量中30个数据作为一组,则所述真实元数据为数组的形式,所述真实元数据包括数据包的长度以及发送所述数据包一端的端口号;

所述联邦学习全局模型的训练方法包括:

获取训练未知网络流量,并提取所述训练未知网络流量中的训练元数据;

将所述训练元数据发送至各个客户端;

从所有客户端中随机选取若干客户端组成客户端集合;

针对客户端集合中的所有客户端,基于所述训练元数据,将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器;

根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器,得到初始全局模型;

对所述初始全局模型进行训练,得到全局模型;

通过所述服务器将所述全局模型发送至各客户端,并迭代执行从所有客户端中随机选取若干客户端组成客户端集合的步骤;

当迭代满足预设条件时,停止迭代,得到联邦学习全局模型;

结合客户端本地二分类器与服务器模型的权重差异以及客户端的本地训练损失选择客户端参与联邦。

2.根据权利要求1所述的基于元数据辅助和联邦学习的未知网络流量分类方法,其特征在于,每个所述客户端的二分类器基于训练得到。

3.根据权利要求1所述的基于元数据辅助和联邦学习的未知网络流量分类方法,其特征在于,每个所述客户端的二分类器均基于预设的损失函数训练得到。

4.根据权利要求1所述的基于元数据辅助和联邦学习的未知网络流量分类方法,其特征在于,所述针对客户端集合中的所有客户端,基于所述训练元数据,将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器包括:

将所述训练元数据输入至每个客户端对应的二分类器,输出与每个二分类器对应的类别概率值;

将类别概率值大于预设概率阈值的二分类器上传至服务器。

5.根据权利要求1所述的基于元数据辅助和联邦学习的未知网络流量分类方法,其特征在于,所述根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器,得到初始全局模型包括:

基于联邦平均算法,将所有成功识别训练未知网络流量的类别的二分类器进行聚合,得到初始全局模型。

6.根据权利要求1所述的基于元数据辅助和联邦学习的未知网络流量分类方法,其特征在于,所述当迭代满足预设条件时,停止迭代,得到联邦学习全局模型包括:

当迭代次数达到预设次数阈值时，停止迭代，得到联邦学习全局模型。

7.一种基于元数据辅助和联邦学习的未知网络流量分类装置，其特征在于，所述装置包括：

获取和提取模块，用于获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；

分类模块，用于将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到；

所述获取和提取模块还用于：

将所述真实未知网络流量中30个数据作为一组，则所述真实元数据为数组的形式，所述真实元数据包括数据包的长度以及发送所述数据包一端的端口号；

所述联邦学习全局模型的训练方法包括：

获取训练未知网络流量，并提取所述训练未知网络流量中的训练元数据；

将所述训练元数据发送至各个客户端；

从所有客户端中随机选取若干客户端组成客户端集合；

针对客户端集合中的所有客户端，基于所述训练元数据，将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器；

根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器，得到初始全局模型；

对所述初始全局模型进行训练，得到全局模型；

通过所述服务器将所述全局模型发送至各客户端，并迭代执行从所有客户端中随机选取若干客户端组成客户端集合的步骤；

当迭代满足预设条件时，停止迭代，得到联邦学习全局模型；

结合客户端本地二分类器与服务器模型的权重差异以及客户端的本地训练损失选择客户端参与联邦。

8.一种智能终端，其特征在于，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如权利要求1-6中任意一项所述的方法。

9.一种非临时性计算机可读存储介质，其特征在于，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如权利要求1-6中任意一项所述的方法。

一种基于元数据辅助和联邦学习的未知网络流量分类方法

技术领域

[0001] 本发明涉及互联网技术领域，尤其涉及的是一种基于元数据辅助和联邦学习的未知网络流量分类方法。

背景技术

[0002] 网络流量分类是将网络流量划分为不同的类别，在网络异常检测、QoS(Quality of Service)、网络监控、流量工程(Traffic Engineering)等网络管理中发挥着重要作用。但是现有技术中进行未知网络流量识别都是采用集中式的分类模型，会出现数据隐私和安全问题，并且对未知网络流量的分类精度不高。

[0003] 因此，现有技术还有待改进和发展。

发明内容

[0004] 本发明要解决的技术问题在于，针对现有技术的上述缺陷，提供一种基于元数据辅助和联邦学习的未知网络流量分类方法，旨在解决现有技术中进行未知网络流量识别都是采用集中式的分类模型，会出现数据隐私和安全问题，并且对未知网络流量的分类精度不高的问题。

[0005] 本发明解决问题所采用的技术方案如下：

[0006] 第一方面，本发明实施例提供一种基于元数据辅助和联邦学习的未知网络流量分类方法，其中，所述方法包括：

[0007] 获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；

[0008] 将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到。

[0009] 在一种实现方式中，所述联邦学习全局模型的训练方法包括：

[0010] 获取训练未知网络流量，并提取所述训练未知网络流量中的训练元数据；

[0011] 将所述训练元数据发送至各个客户端；

[0012] 从所有客户端中随机选取若干客户端组成客户端集合；

[0013] 针对客户端集合中的所有客户端，基于所述训练元数据，将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器；

[0014] 根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器，得到初始全局模型；

[0015] 对所述初始全局模型进行训练，得到全局模型；

[0016] 通过所述服务器将所述全局模型发送至各客户端，并迭代执行从所有客户端中随机选取若干客户端组成客户端集合的步骤；

[0017] 当迭代满足预设条件时，停止迭代，得到联邦学习全局模型。

[0018] 在一种实现方式中，每个所述客户端的二分类器基于训练得到。

[0019] 在一种实现方式中，每个所述客户端的二分类器均基于预设的损失函数训练得到。

[0020] 在一种实现方式中，所述针对客户端集合中的所有客户端，基于所述训练元数据，将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器包括：

[0021] 将所述训练元数据输入至每个客户端对应的二分类器，输出与每个二分类器对应的类别概率值；

[0022] 将类别概率值大于预设概率阈值的二分类器上传至服务器。

[0023] 在一种实现方式中，所述根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器，得到初始全局模型包括：

[0024] 基于联邦平均算法，将所有成功识别训练未知网络流量的类别的二分类器进行聚合，得到初始全局模型。

[0025] 在一种实现方式中，所述当迭代满足预设条件时，停止迭代，得到联邦学习全局模型包括：

[0026] 当迭代次数达到预设次数阈值时，停止迭代，得到联邦学习全局模型。

[0027] 第二方面，本发明实施例还提供一种基于元数据辅助和联邦学习的未知网络流量分类装置，其中，所述装置包括：

[0028] 获取和提取模块，用于获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；

[0029] 分类模块，用于将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到。

[0030] 第三方面，本发明实施例还提供一种智能终端，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如上述任意一项所述的基于元数据辅助和联邦学习的未知网络流量分类方法。

[0031] 第四方面，本发明实施例还提供一种非临时性计算机可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如上述中任意一项所述的基于元数据辅助和联邦学习的未知网络流量分类方法。

[0032] 本发明的有益效果：本发明实施例首先获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；然后将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到；可见，本发明实施例基于元数据和二分类器，根据联邦学习全局模型来对未知网络流量进行分类识别，使得本发明能够一个具有未知网络流量的客户端在保护数据隐私和安全的情况下能从其他客户端学习到未知流量的分类方法。

附图说明

[0033] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本

发明中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0034] 图1为本发明实施例提供的基于元数据辅助和联邦学习的未知网络流量分类方法流程示意图。

[0035] 图2为本发明实施例提供的一种实现方式的基于元数据辅助和联邦学习的未知网络流量分类方法流程示意图。

[0036] 图3为本发明实施例提供的基于元数据辅助和联邦学习的未知网络流量分类装置的原理框图。

[0037] 图4为本发明实施例提供的智能终端的内部结构原理框图。

具体实施方式

[0038] 本发明公开了一种基于元数据辅助和联邦学习的未知网络流量分类方法,为使本发明的目的、技术方案及效果更加清楚、明确,以下参照附图并举实施例对本发明进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0039] 本技术领域技术人员可以理解,除非特意声明,这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是,本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解,当我们称元件被“连接”或“耦接”到另一元件时,它可以直接连接或耦接到其他元件,或者也可以存在中间元件。此外,这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的全部或任一单元和全部组合。

[0040] 本技术领域技术人员可以理解,除非另外定义,这里使用的所有术语(包括技术术语和科学术语),具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是,诸如通用字典中定义的那些术语,应该被理解为具有与现有技术的上下文中的意义一致的意义,并且除非像这里一样被特定定义,否则不会用理想化或过于正式的含义来解释。

[0041] 由于现有技术中,近年来,人们提出了许多流量分类(Traffic Classification)方法来对互联网流量进行分类。这些方法主要分为三类:基于端口的分类方法、基于载荷的分类方法和基于机器学习(Machine Learning)的方法。在传统的流量分类(Traffic Classification TC)问题中,由于存在未知流量以及边缘设备之间互不分享流量信息的问题,边缘设备对于未知流量的分类不能依靠其他边缘设备(已知如何分类该未知流量的边缘设备)的帮助。不仅如此,边缘设备需要将原始流量数据发送到服务器进行集中处理,这不仅会产生大量的通信开销,还会导致隐私泄露和信息安全问题。由于目前数据的隐私问题,客户端之间不能直接分享具体的网络流量中的数据。但是现有技术中进行未知网络流量识别都是采用集中式的分类模型,会出现数据隐私和安全问题,使得未知网络流量的客户端无法从其他客户端学习到未知流量的分类方法。

[0042] 为了解决现有技术的问题,本实施例提供了一种基于元数据辅助和联邦学习的未知网络流量分类方法,上述方法基于元数据和二分类器,根据联邦学习全局模型来对未知网络流量进行分类识别,使得本发明能够一个具有未知网络流量的客户端在保护数据隐私

和安全的情况下能从其他客户端学习到未知流量的分类方法。具体实施时，首先获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；然后将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到。

[0043] 示例性方法

[0044] 本实施例提供一种基于元数据辅助和联邦学习的未知网络流量分类方法，该方法可以应用于互联网技术的智能终端。具体如图1所示，所述方法包括：

[0045] 步骤S100、获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；

[0046] 具体地，真实未知网络流量可以来自于网络，可以是国内网络也可以是国外网络，为了更好的对真实未知网络流量进行分类，本发明实施例提取所述真实未知网络流量中的真实元数据。在本实施例中，可以将真实未知网络流量中30个数据作为一组，则真实元数据为数组的形式。真实元数据可以包括数据包的长度以及发送数据包一端的端口号。因为元数据是对单一应用的整体网络流量的描述，使用元数据(metadata)会大大加快分析未知流量类别的速度。

[0047] 得到真实元数据后，就可以执行如图1所示的如下步骤：S200、将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到。

[0048] 具体地，二分类器(Binary Classifier)在网络流量的元数据(Metadata)上的应用也是一种新的分类网络流量的方式。联邦学习(Federated Learning)是一种新的分布式机器学习框架，可以解决TC中的数据隐私和安全问题。联邦学习是一种新的分布式机器学习框架，它允许多个客户端协作训练全局模型而无需共享原始流量数据。在联邦学习(FL)框架下，原始流量数据保存在本地客户端进行训练，客户端共享学习到的分类模型，而不是原始流量数据，这极大的保护了用户隐私和数据安全。现有的联邦学习方法一般分为四个步骤：首先，服务器向本地客户端广播全局流量分类模型；其次，本地客户端下载全局模型，在本地数据集上使用SVM、Naive Bayes、深度学习等机器学习方法来训练分类模型；在对客户端上的流量数据进行本地分类模型训练后，将本地模型参数上传到服务器；最后，使用FedAvg等聚合算法，对来自客户端的局部模型进行聚合，即对局部模型的权重参数进行加权平均，通过不断的循环迭代上述的四个步骤，最终得到全局模型即为最优模型。为了解决各个客户端的隐私问题，客户端之间不能直接分享具体的网络流量中的数据。而在联邦学习框架下利用元数据的二分类器能在保护用户隐私的同时解决这一问题。具体而言，客户端利用二分类器可以实现对于未知网络流量的分类，而这个分类模型会在客户端训练结束后上传到服务器。服务器会利用这些能识别未知网络流量的类别的二分类器进行聚合，最后会把对未知网络流量的分类信息分享给所有的客户端，也即分享对未知网络流量的分类模型到各个客户端。在本实施例中，结合若干已训练的二分类器和联邦学习得到联邦学习全局模型，再将真实元数据输入至联邦学习全局模型，就可以得到未知网络流量的类别，可以在保护用户隐私的情况下实现高精度的分类，并将分类的类别分享给所有的客户端。基于元数据辅助和联邦学习的未知网络流量分类方法(MEAT)通过结合客户端本地二分类器

与服务器模型的权重差异以及客户端的本地训练损失来选择合适的客户端参与联邦，提高全局模型的预测精度以及收敛速度。

[0049] 在一种实现方式中，所述联邦学习全局模型的训练方法包括如下步骤：获取训练未知网络流量，并提取所述训练未知网络流量中的训练元数据；将所述训练元数据发送至各个客户端；从所有客户端中随机选取若干客户端组成客户端集合；针对客户端集合中的所有客户端，基于所述训练元数据，将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器；根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器，得到初始全局模型；对所述初始全局模型进行训练，得到全局模型；通过所述服务器将所述全局模型发送至各客户端，并迭代执行从所有客户端中随机选取若干客户端组成客户端集合的步骤；当迭代满足预设条件时，停止迭代，得到联邦学习全局模型。

[0050] 具体地，如图2所示，可以通过服务器C_0或者各个客户端$i \in C$在公开的真实网络数据集ISCX2016上下载数据，其中，客户端初始集合：$C=\{C_1,C_2,C_3,\cdots,C_n\}$，可以通过服务器识别出未知网络流量$D_u$，也可以通过客户端识别出未知网络流量$D_u$然后通知服务器，服务器从未知网络流量$D_u$中提取出训练元数据，然后将所述训练元数据发送到客户端初始集合C中的所有客户端，然后在客户端初始集合C中随机选取若干客户端组成新的客户端集合S，这样通过多个客户端同时对未知网络流量的识别，提高联邦学习全局模型收敛速度。针对客户端集合S中的所有客户端i，基于所述训练元数据，将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器。

[0051] 在一种实现方式中，所述针对客户端集合中的所有客户端，基于所述训练元数据，将成功识别训练未知网络流量的类别的客户端对应的二分类器上传至服务器包括如下步骤：将所述训练元数据输入至每个客户端对应的二分类器，输出与每个二分类器对应的类别概率值；将类别概率值大于预设概率阈值的二分类器上传至服务器。

[0052] 具体地，每个客户端对应的二分类器事先经过训练得到，可以对本地客户端的原始网络流量进行分类，得到本地客户端的原始网络流量的类别，分别为未知网络流量和已知网络流量。在本实施例中，每个客户端的二分类器均基于预设的损失函数训练得到。损失函数如下公式：

[0053] $$L_a = -\frac{1}{|D|}\Sigma_{i=1}^{|D|}[p(x_i)log(p(x_i))+(1-p(x_i))log(1-p(x_i))] \quad (1)$$

[0054] 其中，D代表训练数据x的个数。单个训练数据x是一定长度的数组，通常实验中选取30作为数组长度。数组内的元素是训练元数据，例如前面提到的数据包的长度，以及发送数据包一端的端口号。p(xi)代表了预测这个训练元数据xi属于目标流量的概率。这个目标流量就是我们已知的一类流量，标记为1，未知的流量标记为0，0～1就是代表了这个训练数据属于已知流量一类的概率，相应的，1-p(xi)就是xi属于未知流量的概率。

[0055] 得到各个客户端训练好的二分类器后，将训练未知网络流量中所述训练元数据输入至每个客户端对应的二分类器时，会通过所述二分类器输出一个与之对应的类别概率值，当一个二分类器对应的类别概率值大于预设概率阈值(如0.5)时，则表明，该客户端的二分类器能够识别出训练未知网络流量中训练元数据的类别，可以用于对其他客户端训练未知网络流量的类别识别的分享，故将能识别出训练未知网络流量中训练元数据的类别的

二分类器上传至服务器，除此之外，还可以将识别出训练未知网络流量中训练元数据的类别的二分类器的损失信息也发送至服务器端。

[0056] 这样，服务器就接收到多个成功识别未知网络流量的类别的客户端对应的二分类器，然后就可以根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器，得到初始全局模型。相应的，所述根据服务器中接收到的若干成功识别训练未知网络流量的类别的二分类器，得到初始全局模型包括如下步骤：基于联邦平均算法，将所有成功识别训练未知网络流量的类别的二分类器进行聚合，得到初始全局模型。

[0057] 具体地，联邦平均算法(FedAvg算法)：将所有成功识别训练未知网络流量的类别的二分类器的参数进行加权后求平均得到聚合后的参数，采用聚合后的参数更新模型，得到初始全局模型。然后对初始全局模型进行训练，可以基于初始全局模型的损失函数对初始全局模型进行训练，从而得到全局模型，然后通过所述服务器将所述全局模型发送至各个客户端，然后迭代执行从所有客户端中随机选取若干客户端组成客户端集合的步骤，当迭代次数达到预设次数阈值T(可以为20、25、30、35和40等)时，停止迭代，得到联邦学习全局模型。可以通过服务器将联邦学习全局模型分享至各个客户端，各个客户端就可以分享对未知网络流量的分类方法，也就是说，当一个客户端面临一个未知网络流量而无法辨别出该未知网络流量的类别时，会存在其他客户端能够识别出该未知网络流量的类别，客户端可以通过获取其他客户端共享的类别从而得到该未知网络流量的类别。比如，facebook应用的网络流量对于A客户端而言是未知网络流量，A客户端无法识别其类别，但是facebook应用的网络流量对于B客户端而言，其类别是已知的(如聊天)，这样，通过B客户端识别出facebook应用的网络流量的类别后，通过联邦学习的共享，A客户端也就能知道facebook应用的网络流量的类别。

[0058] 本发明在公开的真实网络数据集ISCX2016上进行了实验以证明本方法的优越性。实验结果表明本发明基于元数据辅助和联邦学习的未知网络流量分类方法(MEAT)相比于集中式的分类模型对于未知网络流量的分类，准确度提升了14％。

[0059] 本发明针对未知网络流量分类场景，提出了结合使用元数据的二分类器以及客户端训练损失的联邦学习客户端选择算法MEAT。该算法使得不同客户端对于未知流量的分类模型信息便可以在保留隐私的前提下进行分享，从而保护了用户的隐私，也提高了对于未知流量分类的成功率。同时，通过多个客户端同时对未知网络流量进行识别，使得联邦学习全局模型收敛速度更快。

[0060] 本发明具有以下优点：

[0061] (1)本发明首先提出了将联邦学习框架应用到未知网络流量分类问题上，保护了客户端的数据安全和用户隐私并且减少了由本地数据量不足所导致的模型不准确的影响。

[0062] (2)提出了一种在联邦学习框架下使用元数据辅助和保护隐私的方法：基于元数据辅助和联邦学习的未知网络流量分类方法(MEAT)。这种方法可以使得客户端协同合作，在不共享本地数据的同时，建立一个对未知网络流量的分类共享模型。

[0063] (3)MEAT算法在公开数据集ISCX上对比于传统的集中式模型分类未知数据的方式提升了14.0％的预测准确度。

[0064] 示例性设备

[0065] 如图3中所示，本发明实施例提供一种基于元数据辅助和联邦学习的未知网络流

量分类装置，该装置包括获取和提取模块301和分类模块302，其中：

[0066] 获取和提取模块301，用于获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；

[0067] 分类模块302，用于将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到。

[0068] 基于上述实施例，本发明还提供了一种智能终端，其原理框图可以如图4所示。该智能终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏、温度传感器。其中，该智能终端的处理器用于提供计算和控制能力。该智能终端的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该智能终端的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种基于元数据辅助和联邦学习的未知网络流量分类方法。该智能终端的显示屏可以是液晶显示屏或者电子墨水显示屏，该智能终端的温度传感器是预先在智能终端内部设置，用于检测内部设备的运行温度。

[0069] 本领域技术人员可以理解，图4中的原理图，仅仅是与本发明方案相关的部分结构的框图，并不构成对本发明方案所应用于其上的智能终端的限定，具体的智能终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0070] 在一个实施例中，提供了一种智能终端，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令：获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；

[0071] 将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习全局模型基于若干已训练的二分类器得到。

[0072] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本发明所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0073] 综上所述，本发明公开了一种基于元数据辅助和联邦学习的未知网络流量分类方法，所述方法包括：获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据；其中，所述真实元数据为网络流量数据的部分字节数据；将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别；其中，所述联邦学习

全局模型基于若干已训练的二分类器得到。本发明实施例基于元数据和二分类器，根据联邦学习全局模型来对未知网络流量进行分类识别，使得本发明能够一个具有未知网络流量的客户端在保护数据隐私和安全的情况下能从其他客户端学习到未知流量的分类方法。

[0074] 基于上述实施例，本发明公开了一种基于元数据辅助和联邦学习的未知网络流量分类方法，应当理解的是，本发明的应用不限于上述的举例，对本领域普通技术人员来说，可以根据上述说明加以改进或变换，所有这些改进和变换都应属于本发明所附权利要求的保护范围。

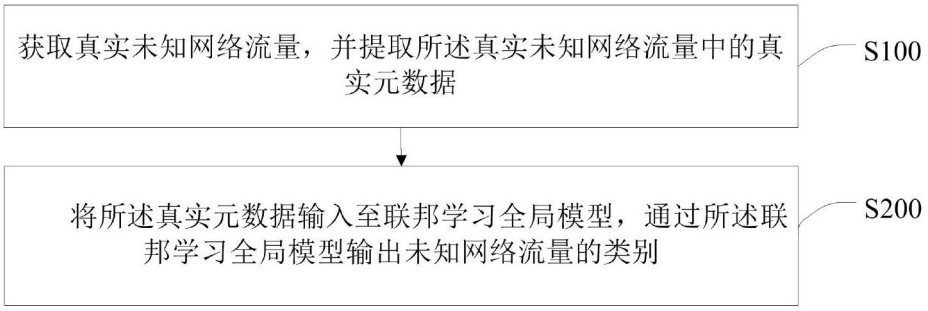

获取真实未知网络流量，并提取所述真实未知网络流量中的真实元数据
S100
将所述真实元数据输入至联邦学习全局模型，通过所述联邦学习全局模型输出未知网络流量的类别
S200

图1

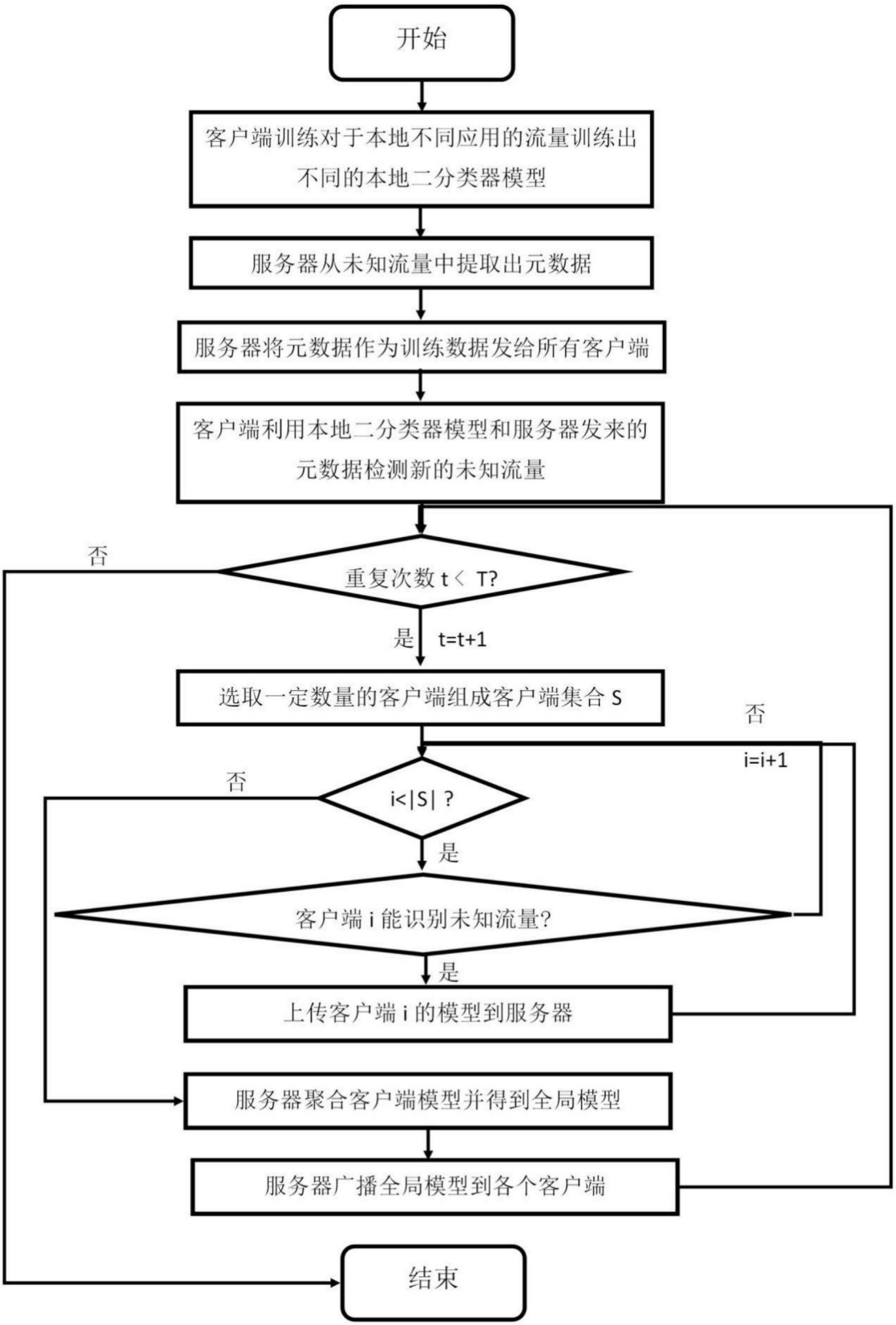

开始
客户端训练对于本地不同应用的流量训练出不同的本地二分类器模型
服务器从未知流量中提取出元数据
服务器将元数据作为训练数据发给所有客户端
客户端利用本地二分类器模型和服务器发来的元数据检测新的未知流量
否
重复次数 t < T?
是
t=t+1
选取一定数量的客户端组成客户端集合 S
否
i=i+1
否
i<|S| ?
是
客户端 i 能识别未知流量?
是
上传客户端 i 的模型到服务器
服务器聚合客户端模型并得到全局模型
服务器广播全局模型到各个客户端
结束

图2


获取和提取模块
301
分类模块
302

图3

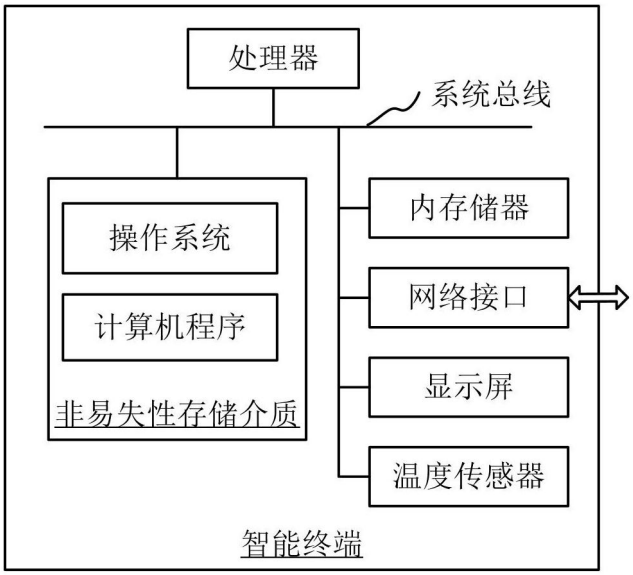

处理器
系统总线
操作系统
计算机程序
非易失性存储介质
内存储器
网络接口
显示屏
温度传感器
智能终端

图4