(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 118869552 B
(45)授权公告日 2025.02.07

(21)申请号 202411346930.2
(22)申请日 2024.09.26
(65)同一申请的已公布的文献号
申请公布号 CN 118869552 A
(43)申请公布日 2024.10.29
(73)专利权人 卓望数码技术(深圳)有限公司
地址 518000 广东省深圳市南山区粤海街道高新区社区高新南七道015号深港产学研基地W601
(72)发明人 郑伟 袁胜
(74)专利代理机构 深圳市恒和大知识产权代理有限公司 44479
专利代理师 林大超
(51)Int.Cl.
H04L 43/0876(2022.01)
H04L 43/04(2022.01)
H04L 43/026(2022.01)
H04L 41/147(2022.01)
H04L 41/16(2022.01)
G06N 3/0442(2023.01)
G06N 3/08(2023.01)

(56)对比文件
CN 116318970 A,2023.06.23
CN 111447190 A,2020.07.24

审查员 钟秀媚

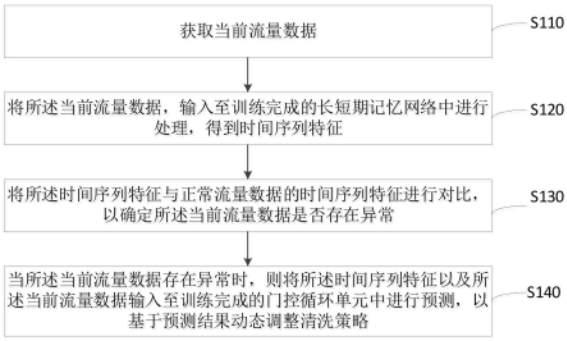
权利要求书2页 说明书12页 附图2页

(54)发明名称

智能流量清洗方法、装置、计算机设备及存储介质

(57)摘要

本申请公开了智能流量清洗方法、装置、设备及介质,其方法,包括:获取当前流量数据;将当前流量数据,输入至训练完成的长短期记忆网络中进行处理,得到时间序列特征,将时间序列特征与正常流量数据的时间序列特征进行对比,以确定当前流量数据是否存在异常;当当前流量数据存在异常时,则将时间序列特征以及当前流量数据输入至训练完成的门控循环单元进行预测,以动态调整清洗策略。采用长短期记忆网络和门控循环单元结合,可捕捉到流量数据中的长期依赖关系,从而更准确地理解正常流量的行为模式。通过实时检测异常流量并自动调整防护策略,可实现毫秒级响应,有效应对各种网络攻击。提高异常检测的准确性,降低误报率和漏报率。



CN 118869552 B

1.一种智能流量清洗方法,其特征在于,所述方法,包括:

获取当前流量数据;

将所述当前流量数据,输入至训练完成的长短期记忆网络中进行处理,得到时间序列特征;

将所述时间序列特征与正常流量数据的时间序列特征进行对比,以确定所述当前流量数据是否存在异常;

当所述当前流量数据存在异常时,将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元,预测得到下一个时间步对应的策略参数;

基于所述策略参数,动态调整清洗策略;

其中,所述将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元之后,包括:

基于所述门控循环单元输出流量感知特征,所述流量感知特征包括隐藏状态、重置门输出以及更新门输出;

基于所述隐藏状态,调整流量清洗力度,其中,所述隐藏状态表征当前流量数据的状态;

基于所述更新门输出,调整流量清洗及时性,其中,所述更新门输出表征当前时间步与上一时间步的结合比例;

基于所述重置门输出,调整流量清洗持久性,其中,所述重置门输出表征上一时间步对应的隐藏状态向当前输入的重置程度。

2.如权利要求1所述的智能流量清洗方法,其特征在于,所述长短期记忆网络,通过如下方式训练得到:

采集历史流量数据,对所述历史流量数据标记正标签与负标签,得到训练样本数据;

构建初始长短期记忆网络;

通过所述训练样本数据,对所述初始长短期记忆网络进行迭代训练;

当符合预设收敛条件时,得到所述训练完成的长短期记忆网络。

3.如权利要求1所述的智能流量清洗方法,其特征在于,所述确定所述当前流量数据是否存在异常,包括:

通过长短期记忆网络,逐个时间步对所述当前流量数据进行处理,得到当前流量数据对应的时间序列特征;

通过所述长短期记忆网络学习并记忆正常流量数据对应的时间序列特征;

将所述当前流量数据对应的时间序列特征与所述正常流量数据对应的时间序列特征进行对比;

当对比不一致,则确定所述当前流量数据存在异常;

其中,所述时间序列特征至少包括流量周期性变化、峰值、谷值以及流量增长趋势中的一个。

4.如权利要求1所述的智能流量清洗方法,其特征在于,所述长短期记忆网络包括长短期记忆网络细胞,所述长短期记忆网络细胞的内部状态通过如下方式更新:

将输入门单元通过遗忘门的方式进行更新,得到更新后的输入门;

基于所述更新后的输入门、输入权重以及遗忘门的循环权重,对所述长短期记忆网络

细胞的内部状态进行更新,其中,所述长短期记忆网络细胞的输出通过输出门进行关闭。

5.如权利要求1所述的智能流量清洗方法,其特征在于,所述方法,还包括:

确定所述门控循环单元在反向传播过程中是否出现异常;

若是,根据预设重力参数与截断阈值,确定是否进行截断处理;

若是,则对截断梯度进行更新。

6.一种智能流量清洗装置,其特征在于,所述装置,包括:

当前流量数据获取单元,用于获取当前流量数据,并进行预处理;

时间序列特征提取单元,用于将所述当前流量数据,输入至训练完成的长短期记忆网络中进行处理,得到时间序列特征;

流量异常确定单元,用于将所述时间序列特征与正常流量数据的时间序列特征进行对比,以确定所述当前流量数据是否存在异常;

清洗策略动态调整单元,用于当所述当前流量数据存在异常时将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元,预测得到下一个时间步对应的策略参数;

基于所述策略参数,动态调整清洗策略;

其中,所述清洗策略动态调整单元,还用于:

基于所述门控循环单元输出流量感知特征,所述流量感知特征包括隐藏状态、重置门输出以及更新门输出;

基于所述隐藏状态,调整流量清洗力度,其中,所述隐藏状态表征当前流量数据的状态;

基于所述更新门输出,调整流量清洗及时性,其中,所述更新门输出表征当前时间步与上一时间步的结合比例;

基于所述重置门输出,调整流量清洗持久性,其中,所述重置门输出表征上一时间步对应的隐藏状态向当前输入的重置程度。

7.一种计算机设备,包括存储器、处理器及存储在所述存储器上并在所述处理器上运行的计算机可读指令,其特征在于,所述处理器执行所述计算机可读指令时实现如权利要求1至5任一项所述智能流量清洗方法。

8.一种可读存储介质,其上存储有计算机可读指令,其特征在于,所述计算机可读指令被处理器执行时实现如权利要求1至5任一项所述智能流量清洗方法。

智能流量清洗方法、装置、计算机设备及存储介质

技术领域

[0001] 本申请涉及网络安全技术领域,尤其涉及一种智能流量清洗方法、装置、计算机设备及存储介质。

背景技术

[0002] 随着网络技术的飞速发展,网络安全问题日益凸显。DDoS攻击等网络威胁频繁出现,严重威胁网络安全。传统的清洗策略主要依赖固定的规则和阈值,难以应对复杂多变的网络攻击。目前人工防护策略实时调整难度大,固定的阈值难以自适应。在大多数抗DDoS防护中,防护策略和关键点的决策主体依旧是人,处置效率很大程度上取决于人工经验,对于没有处置先例的新型攻击事件,易陷入被动应对局面;固定的阈值防护难度大,无法自适应,花费运营人员大量的精力和时间,阈值过大会导致漏防,阈值过小会导致误防,导致防护效果不佳。

发明内容

[0003] 基于此,有必要针对上述技术问题,提供一种智能流量清洗方法、装置、计算机设备及存储介质,以解决上述现有技术中存在的至少一个问题。

[0004] 第一方面,本申请实施例是这样实现的,提供了一种智能流量清洗方法,包括如下步骤:

[0005] 获取当前流量数据;

[0006] 将所述当前流量数据,输入至训练完成的长短期记忆网络中进行处理,得到时间序列特征;

[0007] 将所述时间序列特征与正常流量数据的时间序列特征进行对比,以确定所述当前流量数据是否存在异常;

[0008] 当所述当前流量数据存在异常时,则将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元中进行预测,以基于预测结果动态调整清洗策略。

[0009] 在一实施例中,所述长短期记忆网络,通过如下方式训练得到:

[0010] 采集历史流量数据,对所述历史流量数据标记正标签与负标签,得到训练样本数据;

[0011] 构建初始长短期记忆网络;

[0012] 通过所述训练样本数据,对所述初始长短期记忆网络进行迭代训练;

[0013] 当符合预设收敛条件时,得到所述训练完成的长短期记忆网络。

[0014] 在一实施例中,所述确定所述当前流量数据是否存在异常,包括:

[0015] 通过长短期记忆网络,逐个时间步对所述当前流量数据进行处理,得到当前流量数据对应的时间序列特征;

[0016] 通过所述长短期记忆网络学习并记忆正常流量数据对应的时间序列特征;

[0017] 将所述当前流量数据对应的时间序列特征与所述正常流量数据对应的时间序列

特征进行对比；

[0018] 当对比不一致，则确定所述当前流量数据存在异常；

[0019] 其中，所述时间序列特征至少包括流量周期性变化、峰值、谷值以及流量增长趋势中的一个。

[0020] 在一实施例中，所述长短期记忆网络包括长短期记忆网络细胞，所述长短期记忆网络细胞的内部状态通过如下方式更新：

[0021] 将输入门单元通过遗忘门的方式进行更新，得到更新后的输入门；

[0022] 基于所述更新后的输入门、输入权重以及遗忘门的循环权重，对所述长短期记忆网络细胞的内部状态进行更新，其中，所述长短期记忆网络细胞的输出通过输出门进行关闭。

[0023] 在一实施例中，所述将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元进行预测，以基于预测结果动态调整清洗策略，包括：

[0024] 将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元，预测得到下一个时间步对应的策略参数；

[0025] 基于所述策略参数，动态调整清洗策略。

[0026] 在一实施例中，所述将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元之后，包括：

[0027] 基于所述门控循环单元输出流量感知特征，所述流量感知特征包括隐藏状态、重置门输出以及更新门输出；

[0028] 基于所述隐藏状态，调整流量清洗力度，其中，所述隐藏状态表征当前流量数据的状态；

[0029] 基于所述更新门输出，调整流量清洗及时性，其中，所述更新门输出表征当前时间步与上一时间步的结合比例；

[0030] 基于所述重置门输出，调整流量清洗持久性，其中，所述重置门输出表征上一时间步对应的隐藏状态向当前输入的重置程度。

[0031] 在一实施例中，所述方法，还包括：

[0032] 确定所述门控循环单元在反向传播过程中是否出现异常；

[0033] 若是，根据预设重力参数与截断阈值，确定是否进行截断处理；

[0034] 若是，则对截断梯度进行更新。

[0035] 第二方面，提供了一种智能流量清洗装置，包括：

[0036] 当前流量数据获取单元，用于获取当前流量数据，并进行预处理；

[0037] 时间序列特征提取单元，用于将所述当前流量数据，输入至训练完成的长短期记忆网络中进行处理，得到时间序列特征；

[0038] 流量异常确定单元，用于将所述时间序列特征与正常流量数据的时间序列特征进行对比，以确定所述当前流量数据是否存在异常；

[0039] 清洗策略动态调整单元，用于当所述当前流量数据存在异常时，则将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元中进行预测，以基于预测结果动态调整清洗策略。

[0040] 第三方面，提供了一种计算机设备，包括存储器、处理器及存储在所述存储器上并

在所述处理器上运行的计算机可读指令,所述处理器执行所述计算机可读指令时实现如上述所述智能流量清洗方法。

[0041] 第四方面,提供了一种可读存储介质,其上存储有计算机可读指令,所述计算机可读指令被处理器执行时实现如上述所述智能流量清洗方法。

[0042] 上述智能流量清洗方法、装置、计算机设备及存储介质,其方法实现,包括:获取当前流量数据;将所述当前流量数据,输入至训练完成的长短期记忆网络中进行处理,得到时间序列特征;将所述时间序列特征与正常流量数据的时间序列特征进行对比,以确定所述当前流量数据是否存在异常;当所述当前流量数据存在异常时,则当所述当前流量数据存在异常时,则将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元中进行预测,以基于预测结果动态调整清洗策略。本申请实施例中,采用长短期记忆网络和门控循环单元结合使用的深度学习模型,能够捕捉到流量数据中的长期依赖关系,从而更准确地理解正常流量的行为模式。通过实时检测异常流量并自动调整防护策略,能够实现毫秒级响应,有效应对各种网络攻击。提高异常检测的准确性,降低误报率和漏报率。

附图说明

[0043] 为了更清楚地说明本申请实施例的技术方案,下面将对本申请实施例的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申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0044] 图1是本申请一实施例中智能流量清洗方法的一流程示意图;

[0045] 图2是本申请一实施例中深度学习模型的一网络架构示意图;

[0046] 图3是本申请一实施例中智能流量清洗装置的一结构示意图;

[0047] 图4是本申请一实施例中计算机设备的一示意图。

具体实施方式

[0048] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0049] 在一实施例中,如图1所示,提供一种智能流量清洗方法,包括如下步骤:

[0050] 在步骤S110中,获取当前流量数据;

[0051] 在本申请实施例中,可配置多个代理服务器,并分别按照流量种类、流量格式等,实时对网络流量进行分区域采集,或者可通过日志、访问记录、浏览记录等进行获取。其中,该流量种类具体可包括泛流量、精准流量、公域流量、私域流量和推送流量。泛流量包括微博、新闻频道浏览的信息产生的流量;精准流量是用户通过索引或者关键字进行搜索而产生的具体流量;公域流量是在浏览电商平台、知乎、豆瓣等固定平台等产生的同有流量;私域流量是用户软件中自发性产生的流量,例如群、发表的评论、发表的隐私的保存文件保存的流量;推送流量是固定平台、网页、APP等自动推送用户的数据信息进而产出推送流量。

[0052] 在本申请实施例中,当获取到当前流量数据后,可对流量数据进行预处理,以便用

于模型的预测。示例性地,可去除异常值、重复值或无关数据。从原始网络流量数据中提取关键特征,如包长度、协议类型、源IP和目标IP等。

[0053] 需要说明的是,当用于对模型进行训练以及评估时,则可采集历史流量数据,并将不同类型的历史流量数据样本进行标签化,例如,正常流量标记为0,恶意流量标记为1。

[0054] 在步骤S120中,将所述当前流量数据,输入至训练完成的长短期记忆网络中进行处理,得到时间序列特征;

[0055] 在本申请实施例中,可构建长短期记忆网络和门控循环单元结合使用的深度学习模型,作为清洗策略预测模型。其中,长短期记忆网络因其独特的门控机制和细胞状态设计,能够有效地捕捉并处理序列数据中的长期依赖关系,特别适合处理网络流量这类具有时间序列特性的数据。通过长短期记忆网络模型,可以准确地提取网络流量数据中的关键特征,为后续的智能策略调整提供准确的数据基础。该门控循环单元可基于长短期记忆网络提取的时间序列特征进行流量清洗策略的调整,通过门控循环单元模型能有效缓解梯度消失问题,适合于进行实时策略调整。通过门控循环单元,可以根据网络流量的实时变化,动态调整清洗策略,确保网络流量的高效、安全处理。

[0056] 其中,该时间序列特征包括当前流量数据的周期性变化、峰值和谷值、以及增长趋势中的一个或其任意组合。

[0057] 在步骤S130中,将所述时间序列特征与正常流量数据的时间序列特征进行对比,以确定所述当前流量数据是否存在异常;

[0058] 在本申请实施例中,可采集历史流量数据,并对历史流量数据进行预处理。例如,去除异常值、重复值或无关数据。然后,提取关键特征,例如长度、协议类型、源IP和目标IP等。将不同类型的流量样本进行标签化,例如,正常流量标记为0,恶意流量标记为1,以构建流量样本数据集,便于模型训练和评估。构建初始长短期记忆网络,然后通过流量样本数据集,对该初始长短期记忆网络进行迭代训练,该长短期记忆网络能够捕捉到流量数据中的时间序列特性,从而更准确地理解正常流量的行为模式。

[0059] 当获取当前流量数据后,可将其进行预处理后,并输入至训练完成的长短期记忆网络中进行预测,以得到当前流量数据的时间序列特征,然后将其与训练阶段学习到的正常流量数据的时间序列特征进行对比,如果对比一致,则表示当前流量数据为正常流量,否则,其为异常流量。可以理解的,该异常流量的确定过程可在该长短期记忆网络中进行。

[0060] 在步骤S140中,当所述当前流量数据存在异常时,则将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元中进行预测,以基于预测结果动态调整清洗策略。

[0061] 在本申请实施例中,若检测出当前流量数据存在异常时,可将提取的当前流量数据的时间序列特征以及当前流量数据输入至训练完成的门控循环单元中,通过门控循环单元预测下一个时间步的策略参数值,从而实现对流量清洗策略的动态调整。

[0062] 其中,该策略参数值具体可包括流量阈值、带宽分配规则、优先级、安全告警响应措施等清洗策略规则。基于该清洗策略规则可对当前的清洗策略进行调整,例如限速、全局防护等。

[0063] 其中,该门控循环单元针对当前流量的状态可配置有多种防护措施,例如NS限速、DNS状态防护、HTTP限速、HTTP URI防护、全局防护配置等。

[0064] 本申请实施例,提供了一种智能流量清洗方法,包括:获取当前流量数据;将所述当前流量数据,输入至训练完成的长短期记忆网络中进行处理,得到时间序列特征;将所述时间序列特征与正常流量数据的时间序列特征进行对比,以确定所述当前流量数据是否存在异常;当所述当前流量数据存在异常时,则当所述当前流量数据存在异常时,则将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元中进行预测,以基于预测结果动态调整清洗策略。本申请实施例中,采用长短期记忆网络和门控循环单元结合使用的深度学习模型,能够捕捉到流量数据中的长期依赖关系,从而更准确地理解正常流量的行为模式。通过实时检测异常流量并自动调整防护策略,能够实现毫秒级响应,有效应对各种网络攻击。提高异常检测的准确性,降低误报率和漏报率。

[0065] 在本申请一实施例中,所述长短期记忆网络,可通过如下方式训练得到:

[0066] 采集历史流量数据,对所述历史流量数据标记正标签与负标签,得到训练样本数据;

[0067] 构建初始长短期记忆网络;

[0068] 通过所述训练样本数据,对所述初始长短期记忆网络进行迭代训练;

[0069] 当符合预设收敛条件时,得到所述训练完成的长短期记忆网络。

[0070] 具体地,可采集历史流量数据,并对历史流量数据进行预处理。例如,去除异常值、重复值或无关数据,然后,提取关键特征,例如长度、协议类型、源IP和目标IP等。将不同类型的流量样本进行标签化,例如,正常流量标记为0,恶意流量标记为1,以构建流量样本数据集,便于模型训练和评估。构建初始长短期记忆网络,然后通过流量样本数据集,对该初始长短期记忆网络进行迭代训练,当符合预设收敛条件时,例如迭代次数达到预设次数,例如1000次,或者通过预设损失函数,如交叉熵、正则化计算预测结果的损失值,当损失值小于预设阈值,则表示训练完成,可得到训练完成的长短期记忆网络。

[0071] 参见图2,提供了一种深度学习模型的网络架构,在训练过程中,长短期记忆网络可逐个时间步处理历史流量数据,并通过其内部的记忆机制和门控结构,学习并记忆正常流量数据中的时间序列特性。具体算法公式可如下所示:

[0072] $$f_i^{(t)} = \sigma(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)})$$;

[0073] 其中,$f_i^{(t)}$表示遗忘门;t表示时刻;i表示细胞,由sigmoid(激活函数)单元将权重设置为0和1之间的值;$x_j^{(t)}$表示当前输入向量;$h_j^{(t-1)}$表示当前隐藏层向量,包含所有长短期记忆细胞的输出,b^f、U^f、W^f分别表示偏置、输入权重和遗忘门的循环权重,σ表示sigmoid(激活函数),j表示变量。

[0074] 在本申请一实施例中,所述长短期记忆网络包括长短期记忆网络细胞,所述长短期记忆网络细胞的内部状态通过如下方式更新:

[0075] 将输入门单元通过遗忘门的方式进行更新,得到更新后的输入门;

[0076] 基于所述更新后的输入门、输入权重以及遗忘门的循环权重,对所述长短期记忆网络细胞的内部状态进行更新,其中,所述长短期记忆网络细胞的输出通过输出门进行关闭。

[0077] 具体地,细胞状态是长短期记忆网络的核心,可维持和传递长期依赖关系,使得长

短期记忆网络可在处理时间序列数据时可记住并利用之前学习的历史流量数据的时间序列数据。且细胞状态的更新可使得权重在向下传播过程中不会减弱,有利于进行有效信息的提取。更新长短期记忆网络细胞的内部状态,具体可通过如下公式实现:

[0078] $$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)})$$;

[0079] 其中,$f_i^{(t)}$表示遗忘门;t表示时刻;i表示细胞,由sigmoid(激活函数)单元将权重设置为0和1之间的值;$x_j^{(t)}$表示当前输入向量;$h_j^{(t-1)}$表示当前隐藏层向量,包含所有长短期记忆细胞的输出,b、U、W分别表示偏置、输入权重和遗忘门的循环权重,σ表示sigmoid(激活函数),$s_i^{(t)}$表示当前时间步的细胞状态,$s_i^{(t-1)}$表示上一时间步的细胞状态。

[0080] 其中,外部输入门单元$g_i^{(t)}$可以遗忘门的方式进行更新,使用sigmoid(激活函数)获得一个0和1之间的值,则外部输入门的算法公式可如下所示:

[0081] $$g_i^{(t)} = \sigma(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)})$$;

[0082] 其中,$f_i^{(t)}$表示遗忘门;t表示时刻;i表示细胞,由sigmoid(激活函数)单元将权重设置为0和1之间的值;$x_j^{(t)}$表示当前输入向量;$h_j^{(t-1)}$表示当前隐藏层向量,包含所有长短期记忆细胞的输出,b^g、U^g、W^g分别表示偏置、输入权重和遗忘门的循环权重。

[0083] 此时,长短期记忆网络的细胞的输出$h_i^{(t)}$可由输出门$q_i^{(t)}$关闭,具体可通过如下公式计算得到:

[0084] $$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)}$$;

[0085] $$q_i^{(t)} = \sigma(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)})$$;

[0086] 其中,b^o、U^o、W^o分别表示偏置、输入权重和遗忘门的循环权重,在这些变体中,可以选择使用细胞状态$s_i^{(t)}$作为额外的输入。$s_i^{(t)}$表示当前时间步的细胞状态。$\mathbf{x}^{(t)}$表示代表当前时间点的网络流量数据,包括包大小、传输速度、协议类型等特征。$h^{(t-1)}$表示之前的网络流量状态,$\mathbf{s}^{(t)}$表示长期记忆,可帮助模型理解流量的时间序列特性。$f_i^{(t)}$、$g_i^{(t)}$、$h_i^{(t)}$分别控制信息的遗忘、输入和输出,从而帮助模型适应网络流量的动态变化。$q_i^{(t)}$可作为模型的输出,可表示当前网络流量的正常模式,用于后续的异常检测。tanh表示双曲正切函数。

[0087] 可以理解的,通过上述参数与公式,使得长短期记忆网络可学习并理解网络流量的正常行为模式,并未后续的异常检测和清洗策略提供基础,在实际使用中,若当前流量数据与模型学习到的正常流量模式不符合时,可判定为异常流量,并可采用相应的清洗策略。

[0088] 需要说明的是,输入门决定了在当前时间步中,有多少新的信息将被添加到细胞状态中。输出门控制了从细胞状态到输出的信息量。细胞状态作为额外的输入,实际上是指细胞状态在每个时间步都会被更新,更新后的细胞状态不仅影响当前时间步的输出,还会作为下一个时间步的输入的一部分,从而实现了信息的长期保存和流动。

[0089] 在本申请一实施例中,所述确定所述当前流量数据是否存在异常,包括:

[0090] 通过长短期记忆网络,逐个时间步对所述当前流量数据进行处理,得到当前流量数据对应的时间序列特征;

[0091] 通过所述长短期记忆网络学习并记忆正常流量数据对应的时间序列特征;

[0092] 将所述当前流量数据对应的时间序列特征与所述正常流量数据对应的时间序列特征进行对比;

[0093] 当对比不一致,则确定所述当前流量数据存在异常;

[0094] 其中,时间序列特征至少包括流量周期性变化、峰值、谷值以及流量增长趋势中的一个。

[0095] 具体地,在长短期记忆网络训练阶段,可采用历史流量数据,并对历史流量数据进行标签化,例如,正常流量标记为0,恶意流量标记为1。然后通过标注有标签的历史流量数据对长短期记忆网络进行迭代训练,以通过长短期记忆网络学习正常网络模式的时间序列特征。当获取当前流量数据后,可将其进行预处理后,并输入至训练完成的长短期记忆网络中进行预测,以得到当前流量数据的时间序列特征,然后将其与训练阶段学习到的正常流量数据的时间序列特征进行对比,如果对比一致,则表示当前流量数据为正常流量,否则,其为异常流量。

[0096] 在本申请一实施例中,所述将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元进行预测,以基于预测结果动态调整清洗策略,包括:

[0097] 将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元,预测得到下一个时间步对应的策略参数;

[0098] 基于所述策略参数,动态调整清洗策略。

[0099] 其中,策略参数集合中可包括多个提前配置的用于应对异常流量的策略参数,该策略参数具体可包括流量阈值、带宽分配规则、优先级、安全告警响应措施如阻断、限速或重定向等。

[0100] 具体地,可通过长短期记忆网络提取出当前流量数据的关键特征,即时间序列特征,其具体可包括流量的周期性变化、流量的峰值和谷值、流量的增长趋势等。然后,将提取的关键特征与获取的当前流量数据,以及预先配置的策略参数集合输入至门控循环单元中,通过门控单元进行策略参数的动态调整,可输出得到下一时间步对应的策略参数值,然后可使用该策略参数值进行流量清洗。

[0101] 其中,防护措施具体可包括DNS限速、DNS状态防护、HTTP限速、HTTP URI防护、全局防护配置等。

[0102] 在本申请一实施例中,所述将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元之后,包括:

[0103] 基于所述门控循环单元输出流量感知特征,所述流量感知特征包括隐藏状态、重置门输出以及更新门输出;

[0104] 基于所述隐藏状态,调整流量清洗力度,其中,所述隐藏状态表征当前流量数据的状态;

[0105] 基于所述更新门输出,调整流量清洗及时性,其中,所述更新门输出表征当前时间步与上一时间步的结合比例;

[0106] 基于所述重置门输出,调整流量清洗持久性,其中,所述重置门输出表征上一时间步对应的隐藏状态向当前输入的重置程度。

[0107] 具体地,可通过单个门控单元同时控制遗忘因子以及更新状态单元,具体可通过如下公式实现:

[0108] $h_i^{(t)}=u_i^{(t-1)}h_i^{(t-1)}+(1-u_i^{(t-1)})\sigma(b_i+\sum_j U_{i,j}x_j^{(t)}+\sum_j W_{i,j}r_j^{(t-1)}h_j^{(t-1)})$;

[0109] 其中,u代表更新门,r表示复位门。则u以及r可通过如下公式计算得到:

[0110] $u_i^{(t)}=\sigma(b_i^u+\sum_j U_{i,j}^u x_j^{(t)}+\sum_j W_{i,j}^u h_j^{(t)})$;

[0111] $r_i^{(t)}=\sigma(b_i^r+\sum_j U_{i,j}^r x_j^{(t)}+\sum_j W_{i,j}^r h_j^{(t)})$;

[0112] 其中,b、U、W分别表示偏置、输入权重和遗忘门的循环权重。

[0113] 其中,$h_j^{(t)}$(隐藏状态)表示当前流量数据的状态,其值的大小和变化趋势可以作为调整清洗策略的依据。例如,当$h_j^{(t)}$的值持续增大时,可能意味着异常流量在增加,此时应加大清洗力度。

[0114] 其中,$u_i^{(t)}$(更新门输出),可反映当前时间步的信息与上一时间步信息的结合比例。在清洗策略中,可以将其视为对异常流量变化速度的感知,从而调整清洗的及时性。

[0115] 其中,$r_i^{(t)}$(重置门输出)控制了上一时刻状态向当前输入的重置程度。在清洗策略中,可以将其视为对异常流量持续性的感知,从而调整清洗的持久性。

[0116] 需要说明的是,更新门u可以线性门控任意维度,根据需求将它复制或完全由新的目标状态值替换。复位门r控制当前状态中哪些部分用于计算下一个目标状态,在过去状态和未来状态之间引入了附加的非线性效应。

[0117] 在本申请一实施例中,所述方法,还包括:

[0118] 确定所述门控循环单元在反向传播过程中是否出现异常;

[0119] 若是,根据预设重力参数与截断阈值,确定是否进行截断处理;

[0120] 若是,则对截断梯度进行更新。

[0121] 具体地,为了解决深度神经网络训练过程中的梯度消失或梯度爆炸问题。可采用截断梯度算法,通过截断过大的梯度或提升过小的梯度来防止梯度在反向传播过程中出现异常。通过截断梯度法在截断过程中引入了一个重力参数,用于调节模型系数的稀疏性,使得截断过程更加平滑。当梯度超过设定阈值时,不是直接截断,而是根据重力参数逐渐调整梯度值,使其趋近于阈值。

[0122] 在截断梯度法的更新过程中,重力参数与截断阈值共同决定了系数是否被截断。较大的重力参数会导致更多的系数被截断,从而增加模型的稀疏性。

[0123] 其中,截断梯度更新公式可表示为:

[0124] $$\text{TruncatedGradient}_i=\begin{cases} g & \text{if } \|g\|_2\le\theta \\ \frac{g\cdot\theta}{\|g\|_2}+\frac{\alpha(v-\theta)}{v}\cdot\frac{g}{\|g\|_2} & \text{if } \|g\|_2\in(\theta,v] \\ \frac{g\cdot\alpha}{\|g\|_2} & \text{if } \|g\|_2>v \end{cases};$$

[0125] 其中，$\text{TruncatedGradient}_t$ 表示截断梯度算法，g 表示梯度向量，用于指导参数更新的方向和大小。$\| g \|_2 = \sqrt{\sum_{i=1}^{n} g_i^2}$ 表示梯度范数，g_i 是梯度向量的各个分量。v表示梯度的范数上界，即梯度范数的最大允许值，超过这个值的梯度将被更严厉地截断。α表示重力参数，用于表示影响梯度截断程度的因子。当梯度范数超过截断阈值或范数上界时，重力参数决定了梯度被截断后的保留程度。较小的α值会导致更多的截断，而较大的值则保留更多的原始梯度信息。θ表示截断阈值，当梯度范数超过此阈值时，梯度将被截断，以防止梯度过大导致的训练不稳定。

[0126] 需要说明的是，在训练过程中，上述参数的梯度可能会被截断，以确保训练的稳定性和收敛速度。截断梯度有助于防止因为梯度过大而导致的模型参数更新过快，从而避免了模型训练过程中的震荡或发散。通过使用截断梯度算法优化门控循环单元模型的方法。这种优化方法能够显著提高模型的训练效率和稳定性，从而在实际应用中更好地处理复杂的网络流量数据。

[0127] 在本申请实施例中，采用长短期记忆网络模型对网络流量数据进行深度处理。长短期记忆网络模型因其独特的门控机制和细胞状态设计，能够有效地捕捉并处理序列数据中的长期依赖关系，特别适合处理网络流量这类具有时间序列特性的数据。通过长短期记忆网络模型，可以准确地提取网络流量数据中的关键特征，为后续的智能调整提供准确的数据基础。在提取了网络流量的关键特征后，进一步采用门控循环单元模型进行智能调整流量清洗策略。门控循环单元模型能有效缓解梯度消失问题，适合于进行实时策略调整。通过门控循环单元，可以根据网络流量的实时变化，动态调整清洗策略，确保网络流量的高效、安全处理。并且通过采用截断梯度算法对门控循环单元模型进行优化，有效防止梯度爆炸：通过截断过大的梯度，有效防止了梯度在反向传播过程中无限制地增长，从而避免了梯度爆炸的问题。提高训练稳定性：截断梯度有助于保持训练过程的稳定性，减少模型参数更新的剧烈波动。加快收敛速度：通过控制梯度的大小，截断梯度算法可以帮助优化器更快地找到合适的参数更新方向，从而加速模型的收敛。减少过拟合风险：过大的梯度可能导致模型过度拟合训练数据中的噪声。通过截断梯度，可以减少这种过拟合的风险，提高模型的泛化能力。

[0128] 应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

[0129] 在一实施例中，提供一种智能流量清洗装置，该智能流量清洗装置与上述实施例中智能流量清洗方法一一对应。如图3所示，该智能流量清洗装置包括当前流量数据获取单元10、时间序列特征提取单元20、流量异常确定单元30和清洗策略动态调整单元40。各功能模块详细说明如下：

[0130] 当前流量数据获取单元10，用于获取当前流量数据，并进行预处理；

[0131] 时间序列特征提取单元20，用于将所述当前流量数据，输入至训练完成的长短期记忆网络中进行处理，得到时间序列特征；

[0132] 流量异常确定单元30，用于将所述时间序列特征与正常流量数据的时间序列特征进行对比，以确定所述当前流量数据是否存在异常；

[0133] 清洗策略动态调整单元40,用于当所述当前流量数据存在异常时,则将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元中进行预测,以基于预测结果动态调整清洗策略。

[0134] 在本申请一实施例中,所述长短期记忆网络,通过如下方式训练得到:

[0135] 采集历史流量数据,对所述历史流量数据标记正标签与负标签,得到训练样本数据;

[0136] 构建初始长短期记忆网络;

[0137] 通过所述训练样本数据,对所述初始长短期记忆网络进行迭代训练;

[0138] 当符合预设收敛条件时,得到所述训练完成的长短期记忆网络。

[0139] 在本申请一实施例中,流量异常确定单元30,还用于:

[0140] 通过长短期记忆网络,逐个时间步对所述当前流量数据进行处理,得到当前流量数据对应的时间序列特征;

[0141] 通过所述长短期记忆网络学习并记忆正常流量数据对应的时间序列特征;

[0142] 将所述当前流量数据对应的时间序列特征与所述正常流量数据对应的时间序列特征进行对比;

[0143] 当对比不一致,则确定所述当前流量数据存在异常;

[0144] 其中,时间序列特征至少包括流量周期性变化、峰值、谷值以及流量增长趋势中的一个。

[0145] 在本申请一实施例中,所述长短期记忆网络包括长短期记忆网络细胞,所述长短期记忆网络细胞的内部状态通过如下方式更新:

[0146] 将输入门单元通过遗忘门的方式进行更新,得到更新后的输入门;

[0147] 基于所述更新后的输入门、输入权重以及遗忘门的循环权重,对所述长短期记忆网络细胞的内部状态进行更新,其中,所述长短期记忆网络细胞的输出通过输出门进行关闭。

[0148] 在本申请一实施例中,清洗策略动态调整单元40,还用于:

[0149] 将所述时间序列特征、所述当前流量数据以及策略参数集合输入至训练完成的门控循环单元,预测得到下一个时间步对应的策略参数;

[0150] 基于所述策略参数,动态调整清洗策略。

[0151] 在本申请一实施例中,清洗策略动态调整单元40,还用于:

[0152] 基于所述门控循环单元输出流量感知特征,所述流量感知特征包括隐藏状态、重置门输出以及更新门输出;

[0153] 基于所述隐藏状态,调整流量清洗力度,其中,所述隐藏状态表征当前流量数据的状态;

[0154] 基于所述更新门输出,调整流量清洗及时性,其中,所述更新门输出表征当前时间步与上一时间步的结合比例;

[0155] 基于所述重置门输出,调整流量清洗持久性,其中,所述重置门输出表征上一时间步对应的隐藏状态向当前输入的重置程度。

[0156] 在本申请一实施例中,该装置还包括模型优化单元,用于:

[0157] 确定所述门控循环单元在反向传播过程中是否出现异常;

[0158] 若是,根据预设重力参数与截断阈值,确定是否进行截断处理;

[0159] 若是,则对截断梯度进行更新。

[0160] 在本申请实施例中,采用长短期记忆网络模型对网络流量数据进行深度处理。长短期记忆网络模型因其独特的门控机制和细胞状态设计,能够有效地捕捉并处理序列数据中的长期依赖关系,特别适合处理网络流量这类具有时间序列特性的数据。通过长短期记忆网络模型,可以准确地提取网络流量数据中的关键特征,为后续的智能调整提供准确的数据基础。在提取了网络流量的关键特征后,进一步采用门控循环单元模型进行智能调整流量清洗策略。门控循环单元模型能有效缓解梯度消失问题,适合于进行实时策略调整。通过门控循环单元,可以根据网络流量的实时变化,动态调整清洗策略,确保网络流量的高效、安全处理。并且通过采用截断梯度算法对门控循环单元模型进行优化,有效防止梯度爆炸:通过截断过大的梯度,有效防止了梯度在反向传播过程中无限制地增长,从而避免了梯度爆炸的问题。提高训练稳定性:截断梯度有助于保持训练过程的稳定性,减少模型参数更新的剧烈波动。加快收敛速度:通过控制梯度的大小,截断梯度算法可以帮助优化器更快地找到合适的参数更新方向,从而加速模型的收敛。减少过拟合风险: 过大的梯度可能导致模型过度拟合训练数据中的噪声。通过截断梯度,可以减少这种过拟合的风险,提高模型的泛化能力。

[0161] 关于智能流量清洗装置的具体限定可以参见上文中对于智能流量清洗方法的限定,在此不再赘述。上述智能流量清洗装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0162] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是终端设备,其内部结构图可以如图4所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括可读存储介质。该可读存储介质存储有计算机可读指令。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种智能流量清洗方法。本实施例所提供的可读存储介质包括非易失性可读存储介质和易失性可读存储介质。

[0163] 在本申请实施例中,提供了一种计算机设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令,所述处理器执行所述计算机可读指令时实现如上述所述智能流量清洗方法的步骤。

[0164] 在申请实施例中,提供了一种可读存储介质,所述可读存储介质存储有计算机可读指令,所述计算机可读指令被处理器执行时实现如上述所述智能流量清洗方法的步骤。

[0165] 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,是可以通过计算机可读指令来指令相关的硬件来完成,所述的计算机可读指令可存储于一非易失性可读取存储介质或易失性可读存储介质中,该计算机可读指令在执行时,可包括如上述各方法的实施例的流程。其中,本申请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。

[0166] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。

[0167] 以上所述实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围,均应包含在本申请的保护范围之内。

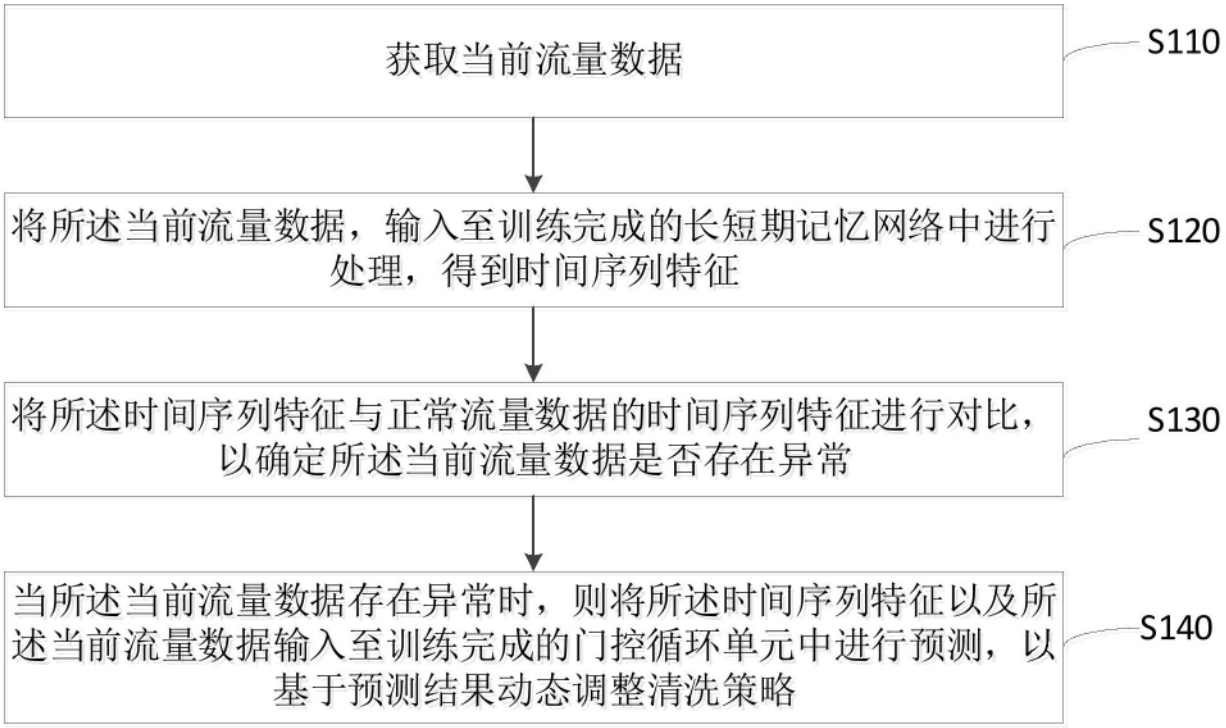

获取当前流量数据
S110
将所述当前流量数据，输入至训练完成的长短期记忆网络中进行处理，得到时间序列特征
S120
将所述时间序列特征与正常流量数据的时间序列特征进行对比，以确定所述当前流量数据是否存在异常
S130
当所述当前流量数据存在异常时，则将所述时间序列特征以及所述当前流量数据输入至训练完成的门控循环单元中进行预测，以基于预测结果动态调整清洗策略
S140

图1

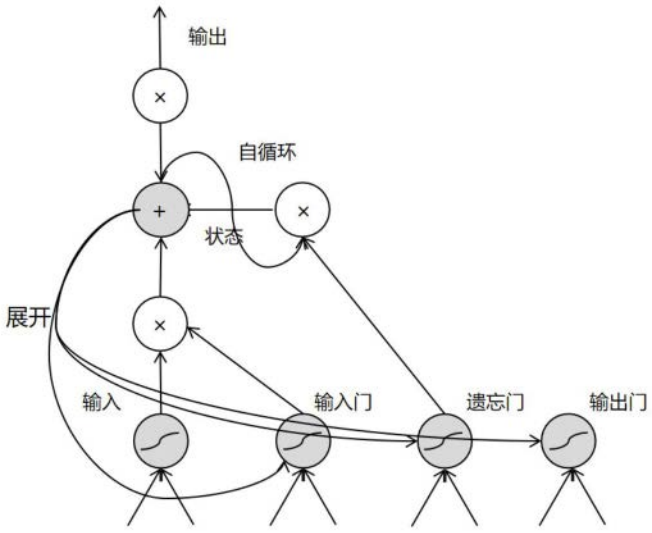

输出
自循环
状态
展开
输入
输入门
遗忘门
输出门

图2


智能流量清洗装置
当前流量数据获取单元
10
时间序列特征提取单元
20
流量异常确定单元
30
清洗策略动态调整单元
40

图3

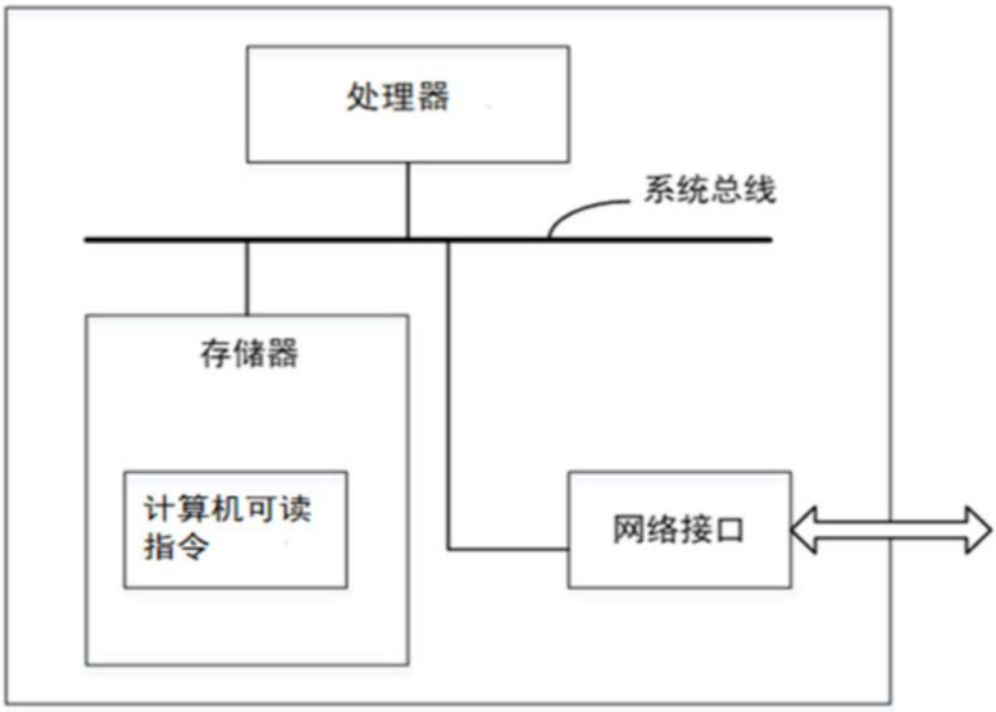

处理器
系统总线
存储器
计算机可读指令
网络接口

图4