(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 112804189 A
(43)申请公布日 2021.05.14

(21)申请号 202011491452.6

(22)申请日 2020.12.17

(71)申请人 北京工业大学
地址 100124 北京市朝阳区平乐园100号

(72)发明人 赖英旭 曹天浩 刘静 王一鹏

(74)专利代理机构 北京思海天达知识产权代理有限公司 11203
代理人 沈波

(51)Int.Cl.
H04L 29/06(2006.01)
H04L 29/08(2006.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)
G06N 20/00(2019.01)

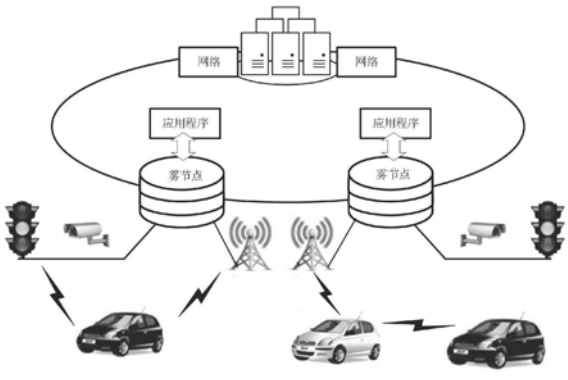
权利要求书2页 说明书6页 附图2页

(54)发明名称

基于云雾协同的车联网入侵检测方法

(57)摘要

本发明公开了基于云雾协同的车联网入侵检测方法,主要由三部分组成,包括:步骤1,由于雾节点和云服务器的计算能力不同,设计了云雾协同防御架构,在资源有限的雾节点将流量数据分为正常数据和可疑数据,在具有强大计算资源的云服务器上将可疑数据具体分类,判别攻击类型。步骤2,由于雾节点资源有限并且网络环境复杂多变的问题,采用CART决策树算法,通过对数据进行检测,能够更快速的确定可疑数据和良性数据。步骤3,针对车联网场景中数据不平衡问题,设计了代价敏感CNN模型,对可疑数据进行具体分类,减少少数攻击漏报率。在模拟现实中的车联网数据集上对算法进行评估,该方法能在较低的资源需求下获得较高的性能。



CN 112804189 A

1.一种基于云雾协同的车联网入侵检测方法,其特征在于:采用云雾协同的数据分类方法,在雾节点采用决策树CART分类器进行粗分类,云服务器采用代价敏感CNN算法进行具体分类,包括:

步骤1,将车联网数据转化为特征向量数据集,特征向量集具体包括802.11p协议IP地址及类型,UDP数据报和IP数据报中的时间、源IP、目的IP、协议名、包大小、端口号、flag信息,以及丢包率、通信链接次数。在资源有限的雾节点利用决策树CART算法对特征向量数据集进行学习,获得决策树CART分类器;

步骤2,在雾节点采用决策树CART将特征向量数据集初步分类,初步分类后的特征向量数据发送至云服务器;

步骤3,在云服务器上,部署代价敏感CNN算法,代价敏感CNN算法对雾节点发送的数据进行具体分类。

2.根据权利要求1所述的一种基于云雾协同的车联网入侵检测方法,其特征在于,所述步骤2中在雾节点采用决策树CART算法进行初步分类,包括:

在车联网的雾节点检测中采用CART决策树算法,CART决策树通过选取GINI系数最小的属性作为根节点的分裂属性,利用二元递归分裂方法形成一种二叉树形式的简单决策树,并且在雾节点进行二分类时效率最高,适合雾节点资源有限并且实时性检测要求。

3.根据权利要求1所述的一种基于云雾协同的车联网入侵检测方法,其特征在于,所述步骤2中在雾节点和云服务器进行不同计算任务分配实现协同,具体协同步骤包括:

步骤21,雾节点将数据进行二分类,分为正常数据和可疑数据。如果雾节点检测到正常数据,则在本地处理,减少发送到云服务器的数据,用于保护智能交通环境下的用户隐私数据。

步骤22,在雾节点上,如果检测到的数据为异常数据,则雾节点将数据发送至云服务器。

步骤23,在云服务器上,代价敏感CNN算法对异常数据进行多分类,得到具体的攻击类型。

步骤24,云服务器中的响应系统会将结果发送至雾节点端的管理员,管理员发现被感染的智能设备,并采取措施实现雾节点与云服务器的协同工作。

4.根据权利要求1所述的一种基于云雾协同的车联网入侵检测方法,其特征在于,所述步骤3中代价敏感CNN在CNN的softmax和loss层之间加入代价矩阵ξ,通过联合优化自动更新参数,具体包括:

步骤31,雾节点筛选出来的可疑数据传递给代价敏感CNN算法,将数据标签更新,更改为具体的攻击标签。为减少类别不平衡对算法的影响,修改CNN的最后一层,在softmax和loss层之间加入一个代价矩阵;

步骤32,在计算分类损失之前,代价矩阵的结果被压缩到[0,1]之间,损失函数采用的是交叉熵损失函数;

步骤33,代价矩阵ξ所添加的位置,在softmax的公式中$\frac{e^{o_n}}{\sum_k e^{o_k}}$,每一个元素的指数值前,都相应的乘上一个代价值,所有的代价值构成代价矩阵的值,其中softmax公式中o_n表示经过两层CNN的输出。

步骤34，使用交叉熵损失函数时，需要更新代价敏感CNN参数θ和代价矩阵参数ξ，采用联合优化方式更新θ和ξ；

5.根据权利要求4所述的步骤34采用联合优化方式更新θ和ξ，具体包括：

步骤51，对于θ的优化，使用误差反向传播的随机梯度下降。为了优化ξ，再次使用梯度下降算法来计算步长的方向来更新参数，具体如下；

步骤52，创建代价敏感CNN网络，初始化神经网络参数θ，将代价矩阵，误差初始化设为1；

步骤53，epoch循环开始，直到达到最大epoch数；

步骤54，计算梯度grad(x,d,F(ξ))，更新梯度参数，其中，x为数据，d为数据标签；

步骤55，batch循环内，前向传播得到输出，反向传播得到梯度，更新网络参数，达到最大batch数，则退出该循环；

步骤56，前向传播得到误差，如果误差大于设定误差，则代价矩阵的学习率缩小100倍，更新误差；

步骤57，epoch循环停止，退出循环；

步骤58，得到代价矩阵参数ξ和学习参数θ最优值。利用待识别的特征向量数据集对代价敏感CNN算法进行训练，得到代价敏感CNN算法分类器。

基于云雾协同的车联网入侵检测方法

技术领域

[0001] 本发明涉及车联网网络安全技术领域,特别涉及一种基于云雾协 同的车联网入侵检测方法。

背景技术

[0002] 智能交通的快速发展使车辆能够与相邻车辆或网络基础设施进 行通信,能够及时获取交通状况,提高安全性和效率,然而这也带来 了很多安全问题。黑客的攻击使得车联网的发展存在重大安全隐患, 攻击者利用漏洞访问网络并篡改机密数据,它会导致更多的事故,改 变安全驾驶的意义。攻击行为可以破坏车联网的系统功能,也可以出 于自己的目的滥用车联网。例如一些黑客通过盗窃车载设备,对车辆 内部网络进行渗透以及利用外部网络实现对车辆的攻击,然后利用这 些被攻击的异常车辆对车联网环境中其他用户进行干扰,这将严重损 害用户的利益甚至是威胁用户的人身安全。

[0003] 车联网是一个动态性很强的快速移动网络,因此车辆之间共享信 息的实时性是非常重要的。由于车辆之间相遇的时间很短,而且对接 收到的信息需要迅速采取行动,因此迅速确定信息的可靠性非常重要。 密码技术涉及成对密钥和开销,涉及计算成本、存储和时间,并且密 钥被盗等行为会导致车联网被入侵,当攻击从车辆内部发起时,更加 难以防范。因此,车联网网络中必须部署入侵检测系统来检测攻击。

[0004] 除了这些与安全相关的挑战之外,车辆还需要处理收集和接收其 他车辆的数据。如果将收集到的交通数据发送到云上执行所需的计算, 然后将结果传达给车辆,这样可以限制车辆之间的计算和通信开销, 提高车辆的私密性。然而,由于道路信息是时间敏感的,这种解决方 案可能是低效的。在雾计算中,雾节点位于终端用户和云之间,利用 路侧单元作为雾节点,雾计算可以作为道路状况计算的替代方法。在 这种情况下,路侧单元从每个路侧单元区域内的车辆收集交通数据, 通过路侧单元对收集到的数据进行分析,提取道路状况。车辆之间的 通信、检测、定位,可以间接地通过雾节点进行交互。

[0005] 因此,为解决车联网的信息安全问题,本文提出基于云雾协同的 车联网入侵检测方法,充分考虑车联网相比原传统互联网的特殊性, 主要是计算能力、存储能力、安全性要求更高等特点,利用当前在诸 多领域都取得突破的机器学习、深度学习技术来构建入侵检测模型。

发明内容

[0006] 本发明所要解决的技术问题是提供一种车联网网络入侵检测的 方法,用于解决高度动态的车联网网络安全问题。本发明解决上述技 术问题的技术方案如下,一种基于云雾协同的车联网入侵检测方法, 包括:

[0007] 步骤1,将车联网数据转化为特征向量数据集,特征向量集具体 包括802.11p协议IP地址及类型,UDP数据报和IP数据报中的时间、源IP、目的IP、协议名、包大小、端口号、flag等信息,以及丢包率、通信链接次数等。在资源有限的雾节点利用决策树CART算法对

该特 征向量数据集进行学习,获得决策树CART分类器;

[0008] 步骤2,在雾节点采用决策树CART将数据初步分类,雾节点将 初步分类结果发送至云服务器;

[0009] 步骤3,在云服务器上,部署代价敏感CNN算法,对雾节点发 送的数据进行具体分类。

[0010] 本发明的关键技术点在于:在车联网的雾节点检测中首次采用 CART决策树算法,该算法具有模型简单和规则提取简单的特点,利 用二元递归分裂方法形成一种二叉树形式的简单决策树,适合雾节点 资源有限并且实时性检测要求;根据雾节点资源有限而云服务器资源 无限的特点,在雾节点和云服务器进行不同的计算任务分配,实现协 同计算:雾节点将数据分成正常数据和可疑数据,将可疑数据发往云 服务器,在云服务器进行具体攻击类别检测;云服务器端采用代价敏 感CNN,即在CNN的softmax和loss层之间加入代价矩阵,并通过 联合优化自动更新参数,提高对攻击的检测准确度。

[0011] 本发明的有益效果是:

[0012] 一、本发明通过引入雾计算,避免了将收集到的交通数据全部发 送到云上执行所需的计算,降低端到端时延,并且采用CART算法检 测经过雾节点的流量,满足车联网的实时性要求。

[0013] 二、本发明采用云雾协同方式,使雾节点和云服务器协同工作, 更好地利用不同设备的存储和计算优势,检测环境中的攻击行为。

[0014] 三、本发明通过改进CNN算法,使其能够更好地处理实际场景 中的不平衡数据,准确地检测到攻击行为,保护云服务器中数据的安 全。

[0015] 综上所述,本发明采用CART算法能够更快的检测出攻击行为, 满足雾节点的实时性要求。采用云雾协同方式,可有效地利用雾节点 和云服务器中的资源。在车联网服务器端,基于代价敏感的CNN方 法可以提高对不平衡数据的检测准确率。本发明能够从车联网的网络 流量中检测出异常行为,保护车联网网络安全。

附图说明

[0016] 图1是本发明的总体结构示意图。

[0017] 图2是本发明云雾协同检测的示意图。

[0018] 图3是本发明采用的CNN算法的模型示意图。

具体实施方式

[0019] 以下结合附图对本发明的原理和特征进行描述,所举实例只用于 解释本发明,并非用于限定本发明的范围。

[0020] 实施例一

[0021] 实施例一采用CART决策树算法,该算法具有模型简单和规则提 取简单的特点,利用二元递归分裂方法形成一种二叉树形式的简单决 策树,适用于雾节点上的入侵检测算法。算法原理如下:

[0022] 步骤1计算属性中各属性的GINI系数,选取GINI系数最小的 属性作为根节点的分裂属性。对于连续属性,需计算其分割阈值,按 分割阈值将其离散化,并计算其GINI系数;

对离散属性,需将样本 集按照该离散属性取值的可能子集进行划分,如该离散属性有N个,则其有效子集有2^n-2个,然后选择GINI系数最小的子集作为该离 散型属性的划分方式,该最小GINI系数作为该离散属性的GINI系 数。

[0023] GINI系数的计算:

[0024] (1)假设整个样本集为S,类别集为{C_1,C_2,...,C_n},总共分为n 类,每个类对应一个样本子集S_i。令|S|为样本集S的样本数,|C_i|为 样本集S中属于类C_i的样本数,则样本集的GINI系数定义如下

[0025] $$Gini(S) = 1 - \sum_{i=1}^{n} p_i^2$$

[0026] 其中,p_i=|C_i|/|S|为样本集中样本属于类C_i的概率。

[0027] (2)在只有二元分裂的时候,对于训练样本集S中的属性A将 S分成的子集S_1和S_2,给定划分S的GINI系数如下公式

[0028] $$Gini_A(S) = \frac{|S_1|}{S} Gini_A(S_1) + \frac{|S_2|}{S} Gini_A(S_2)$$

[0029] 其中,|SK|/|S|为第k个子集占整个样本集的权值。

[0030] 步骤2若分裂属性是连续属性,样本集按照在该属性上的取值, 分成<=T和>T的两部分,T为该连续属性的分割阈值;若分裂属 性是离散属性,样本集按照在该属性上的取值是否包含在该离散属性 具有最小GINI系数的真子集中,分成两部分。

[0031] 步骤3对根节点的分裂属性对应的两个样本子集S_1和S_2,采用 与步骤1相同的方法递归的建立树的子节点。如此循环下去,直到所 有子节点中的样本属于同一类别或没有可以选作分裂属性的属性为 止。

[0032] 步骤4对生成的决策树进行剪枝。

[0033] 基于上述方法,本发明采用机器学习领域中通常使用的检测时间, 召回率(recall)评价指标来对算法的有效性和可靠性进行评价。评价指 标定义如下:

[0034] 检测时间=T2(检测完时间)-T1(开始检测时间)

[0035] $$\text{recall} = \frac{TP}{TP+FN}$$

[0036] 实施例二

[0037] 如图1所示,实施例二是车联网环境下,本发明的总体结构示意 图,主要分为三部分:云服务器,雾节点和终端设备。如图2所示, 为合理利用雾计算与云计算系统中的资源,有效地执行入侵检测任务, 本发明的云雾协同检测方法如下:

[0038] 步骤21,雾节点将数据进行二分类,分为正常数据和可疑数据。如果雾节点检测到正常数据,则在本地处理,减少发送到云服务器的 数据,用于保护智能交通环境下的用户隐私数据。

[0039] 步骤22,在雾节点上,如果检测的数据为异常数据,则雾节点 将异常数据发送至云服务器。

[0040] 步骤23,在云服务器采用代价敏感CNN算法对异常数据进行多 分类,得到具体的

攻击类型。

[0041] 步骤24,云服务器中的响应系统会将结果发送至雾节点端的管 理员,管理员可以发现被感染的智能设备,并采取相应的措施。从而 实现雾节点与云服务器的协同工作。

[0042] 实施例三

[0043] 实施例三是对云服务器上采用的CNN算法的改进。在实际生活 中,通过智能交通雾节点的网络流量中存在大量的正常流量和少量的 异常流量,因此本发明尝试将代价敏感自动学习应用于不平衡数据的 卷积神经网络。

[0044] 步骤1,本发明提出了一个新的代价矩阵ξ,用于修改CNN的最 后一层,在softmax和loss层之间。本发明引入了新的代价矩阵,来 使得算法模型对不频繁的类进行正确分类。因此,根据交叉熵损失函 数F使用代价矩阵ξ修改CNN输出O,如下所示:

[0045] $$y^{(i)} = \mathcal{F}\left(\xi_p,\ O^{(i)}\right):\ y_p^{(i)} \geq y_j^{(i)}\ \ \forall j \neq p \qquad \text{(公式 2)}$$

[0046] 其中$y^{(i)}$表示修改后的输出,p表示期望的类,$\mathcal{F}:R \to R$表示一种 函数,具体为交叉熵损失函数,$O^{(i)}$为CNN的输出,$y_p^{(i)} \geq y_j^{(i)}$表示 修改后的期望的类会比其他类输出值高。

[0047] 步骤2,本方法解决了CNN训练中的类不平衡问题,为此引入 一个代价敏感误差函数,它可以表示为训练集上的平均损失

[0048] $$\mathrm{E}(\theta,\ \xi) = \frac{\sum_{i=1}^{N} \ell\left(d^{(i)}, y^{(i)}{}_{\theta,\ \xi}\right)}{M} \qquad \text{(公式 3)}$$

[0049] 其中,损失层之前的预测输出y受θ和ξ参数影响,θ是CNN参 数,ξ是代价矩阵参数,M是训练集的总数,N表示输出层的神经元 总数,$l(d^{(i)}, y^{(i)}{}_{\theta,\xi})$为交叉熵损失函数,$d \in \{0,1\}^{1\times N}$是期望输出 ($\Sigma_n d_n := 1$),$y^{(i)}$表示求出的softmax值。模型在训练集上表现不佳 时,错误会更大,学习算法的目标是找到最佳参数(θ,ξ),从而降低 代价的平均损失。因此优化目标由

[0050] $(\theta^*,\xi^*) = \mathrm{argmin}\ E(\theta,\xi)$ (公式4)

[0051] 式中的损失函数选择交叉熵损失函数,这种损失最大化了预测与 期望输出的紧密度,交叉熵损失函数如下:

[0052] $l(d,y) = -\Sigma_n(d_n \log\ y_n)$ (公式5)

[0053] d_n是期望输出($\Sigma_n d_n := 1$),y_n表示求出的softmax值。其中y_n依赖于类的代价矩阵,并且与softmax函数的输出o_n相关,如下公式 为代价矩阵添加的位置:

[0054] $$y_n = \frac{\xi_{p,n}\ e^{o_n}}{\sum_k \xi_{p,k}\ e^{o_k}} \qquad \text{(公式 6)}$$

[0055] 其中第n个元素的softmax输出为$\frac{e^{o_n}}{\sum_k e^{o_k}}$,其为第n个元素的指数 与所有元素指数和的比值。

[0056] 步骤3,最优参数学习

[0057] 当使用交叉熵损失函数时,目标是共同学习参数θ和类相关损失 函数参数ξ。对于联合优化,通过保持一个固定的参数并使另一个参 数的代价最小化来交替求解这两种类

型的参数。算法如下：

算法 1. 最优参数学习(θ，ξ)

输入：训练集(x，d)，验证集(x_V，d_V)，训练最大轮次 epochs（M_{ep}），θ的学习率γ_θ，ξ的学习率γ_ξ

输出：学习到的参数(θ^*，ξ^*)

1：创建 CNN 网络
2：初始化网络参数θ（随机初始化）
3：使ξ=1，val-err=1

[0058]

4：for $e \in [1, M_{ep}]$（epoch 的个数）
5：计算梯度 grad(x，d，F(ξ))，得到ξ的梯度$grad_\xi$
6：更新代价参数(ξ，γ_ξ，$grad_\xi$)，得到ξ^*
7：$\xi \leftarrow \xi^*$
8：for $b \in [1, B]$ (batch 的个数)
9：前向传播(x_b，d_b，Net,θ)，得到输出out_b
10：反向传播(out_b，x_b，d_b，Net,θ,ξ)得到梯度$grad_b$
11：更新网络参数(Net,θ,γ_θ，$grad_b$)，得到θ^*
12：$\theta \leftarrow \theta^*$
13：结束 for 循环
14：前向传播（x_V，d_V,Net,θ)，得到$val-err^*$

[0059]

15：如果 $val-err^* > val-err$ ，那么
16：$\gamma_\xi \leftarrow \gamma_\xi * 0.01$
17：$val-err \leftarrow val-err^*$
18：结束
19：结束循环
20：返回(θ^*，ξ^*)

[0060] 本文采用上述方法对CNN算法进行改进，并将改进后的算法应 用到智能交通的云服务器上，实现对入侵数据的分类。

[0061] 步骤4，CNN模型结构如图3所示。模型由两个卷积层、两个池 化层、一个全连接层

和一个dropout层以及softmax分类器组成。

[0062] 基于上述方法,本发明采用机器学习领域中通常使用的检准确率 (precision)评价指标来对算法的有效性和可靠性进行评价。评价指 标定义如下:

[0063] $$precision=\frac{TP}{TP+FP}$$

[0064] 应当理解,虽然本说明书根据实施方式加以描述,但是并非每个 实施方式仅包含一个独立的技术方案,说明书的这种叙述方式仅仅是 为了清楚起见,本领域的技术人员应当将说明书作为一个整体,各个 实施方式中的技术方案也可以适当组合,按照本领域技术人员的理解 来实施。

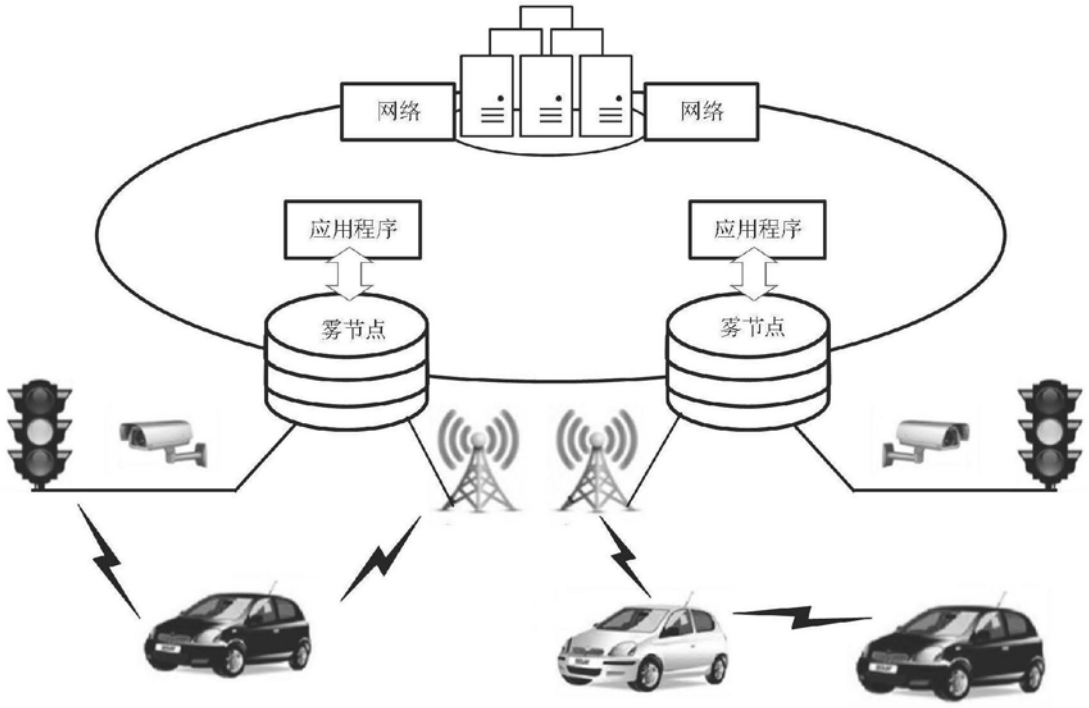

网络
网络
应用程序
应用程序
雾节点
雾节点

图1

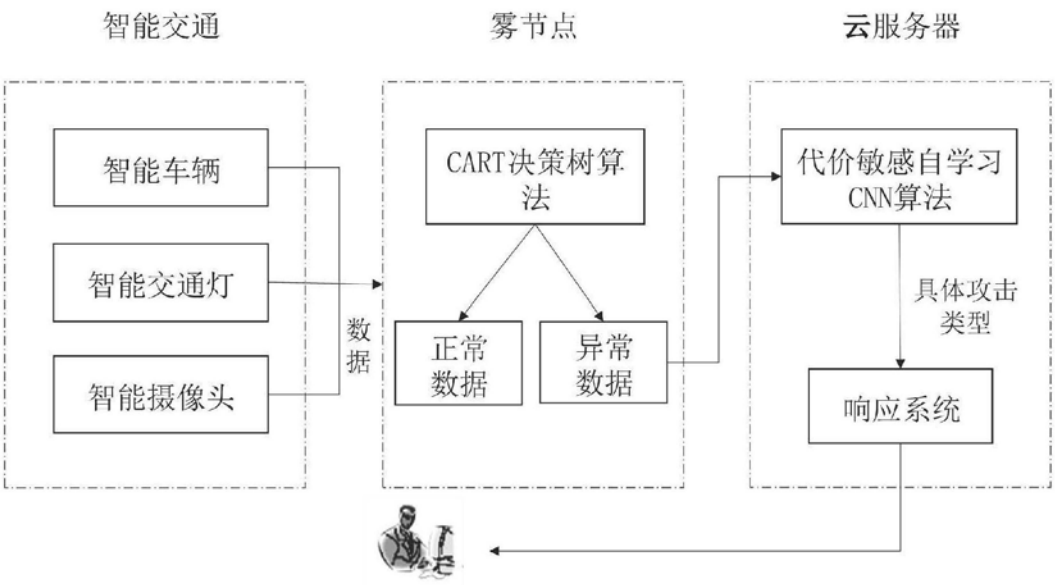

智能交通
雾节点
云服务器
智能车辆
智能交通灯
智能摄像头
数据
CART决策树算法
正常数据
异常数据
代价敏感自学习CNN算法
具体攻击类型
响应系统

图2

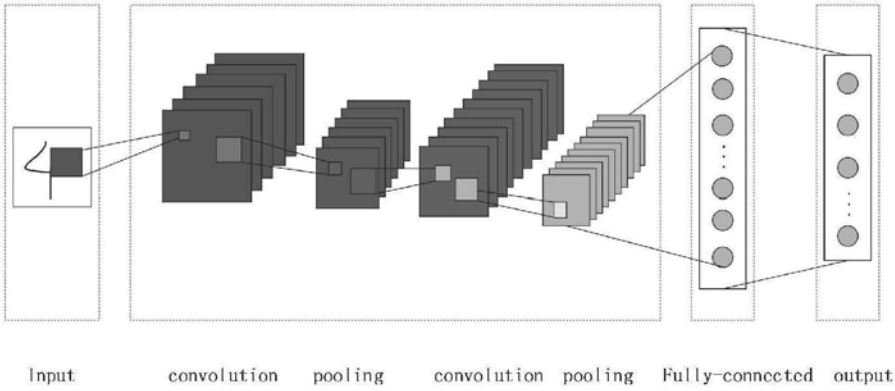

Input
convolution
pooling
convolution
pooling
Fully-connected
output

图3