(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109688030 B
(45)授权公告日 2020.11.03

(21)申请号 201910140315.9
(22)申请日 2019.02.26
(65)同一申请的已公布的文献号
申请公布号 CN 109688030 A
(43)申请公布日 2019.04.26
(73)专利权人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦三层
(72)发明人 刘焱
(74)专利代理机构 北京市铸成律师事务所 11313
代理人 包莉莉 武晨燕
(51)Int.Cl.
H04L 12/26(2006.01)
H04L 12/24(2006.01)
H04L 12/40(2006.01)

审查员 林丽燕

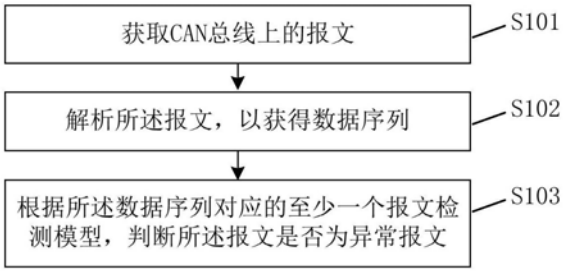
权利要求书3页 说明书10页 附图6页

(54)发明名称

报文检测方法、装置、设备和存储介质

(57)摘要

本发明实施例提出一种报文检测方法、装置、设备和存储介质。该方法包括:获取CAN总线上的报文;解析所述报文,以获得数据序列;根据所述数据序列对应的至少一个报文检测模型,判断所述报文是否为异常报文;其中,所述报文检测模型采用样本报文中的样本数据序列训练得到。本发明实施例的技术方案可以识别出指令级别和参数级别的异常报文,发现未知的入侵行为,还可以避免绕过和漏报,减少因人工维护规则而造成的成本浪费。



CN 109688030 B

1.一种报文检测方法,其特征在于,包括:

获取CAN总线上的报文;

解析所述报文,以获得数据序列;

根据所述数据序列对应的至少一个报文检测模型,判断所述报文是否为异常报文;

其中,所述报文检测模型采用样本报文中的样本数据序列训练得到;

所述报文检测模型包括参数检测模型,根据所述数据序列对应的至少一个报文检测模型,判断所述报文是否为异常报文,包括:

从所述数据序列中确定至少一个字符类型序列,所述字符类型序列包括目标参数中每一位的字符类型,所述目标参数为与所述字符类型序列对应的参数;

根据所述目标参数对应的参数检测模型,判断所述报文是否为异常报文;一个所述参数检测模型对应一个参数。

2.根据权利要求1所述的方法,其特征在于,根据所述目标参数对应的参数检测模型,判断所述报文是否为异常报文,包括:

将所述字符类型序列输入所述目标参数对应的参数检测模型;

获取所述目标参数对应的参数检测模型输出的第一概率;

判断所述第一概率是否小于所述目标参数对应的第一概率阈值;

如果小于所述第一概率阈值,则判定所述报文为异常报文。

3.根据权利要求1所述的方法,其特征在于,所述报文检测模型包括指令检测模型,根据所述数据序列对应的至少一个报文检测模型,判断所述报文是否为异常报文,包括:

从所述数据序列中确定指令序列,所述指令序列包括多条按时间排序的指令;

根据所述指令检测模型,判断所述报文是否为异常报文。

4.根据权利要求3所述的方法,其特征在于,根据所述指令检测模型,判断所述报文是否为异常报文,包括:

将所述指令序列输入所述指令检测模型;

获取所述指令检测模型输出的第二概率;

判断所述第二概率是否小于第二概率阈值;

如果小于所述第二概率阈值,则判定所述报文为异常报文。

5.根据权利要求1所述的方法,其特征在于,还包括:

在车辆正常运行过程中,获取所述车辆的CAN总线上的样本报文;

解析所述样本报文,以获得样本数据序列;

基于所述样本数据序列训练隐式马尔可夫模型,以得到至少一个报文检测模型。

6.根据权利要求5所述的方法,所述报文检测模型包括参数检测模型,基于所述样本数据序列训练隐式马尔可夫模型,以得到至少一个报文检测模型,包括:

从所述样本数据序列中确定多个样本字符类型序列,一个样本字符类型序列对应一个参数,所述样本字符类型序列包括对应参数中每一位的字符类型;

基于各样本字符类型序列分别训练隐式马尔可夫模型,以得到多个参数检测模型,一个参数检测模型对应一个参数。

7.根据权利要求5所述的方法,其特征在于,所述报文检测模型包括指令检测模型,基于所述样本数据序列训练隐式马尔可夫模型,以得到至少一个报文检测模型,包括:

从所述样本数据序列中确定样本指令序列，所述样本指令序列包括多条按时间排序的样本指令；

基于所述样本指令序列训练隐式马尔可夫模型，以得到指令检测模型。

8.一种报文检测装置，其特征在于，包括：

第一获取模块，用于获取CAN总线上的报文；

第一解析模块，用于解析所述报文，以获得数据序列；

判断模块，用于根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文；

其中，所述报文检测模型采用样本报文中的样本数据序列训练得到；

所述报文检测模型包括参数检测模型，所述判断模块包括：

第一确定子模块，用于从所述数据序列中确定至少一个字符类型序列，所述字符类型序列包括目标参数中每一位的字符类型，所述目标参数为与所述字符类型序列对应的参数；

第一判断子模块，用于根据所述目标参数对应的参数检测模型，判断所述报文是否为异常报文；一个所述参数检测模型对应一个参数。

9.根据权利要求8所述的装置，其特征在于，所述第一判断子模块包括：

第一输入单元，用于将所述字符类型序列输入所述目标参数对应的参数检测模型；

第一获取单元，用于获取所述目标参数对应的参数检测模型输出的第一概率；

第一判断单元，用于判断所述第一概率是否小于所述目标参数对应的第一概率阈值；

第一判定单元，用于在所述第一概率小于所述第一概率阈值的情况下，判定所述报文为异常报文。

10.根据权利要求8所述的装置，其特征在于，所述报文检测模型包括指令检测模型，所述判断模块包括：

第二确定子模块，用于从所述数据序列中确定指令序列，所述指令序列包括多条按时间排序的指令；

第二判断子模块，用于根据所述指令检测模型，判断所述报文是否为异常报文。

11.根据权利要求10所述的装置，其特征在于，所述第二判断子模块包括：

第二输入单元，用于将所述指令序列输入所述指令检测模型；

第二获取单元，用于获取所述指令检测模型输出的第二概率；

第二判断单元，用于判断所述第二概率是否小于第二概率阈值；

第二判定单元，用于在所述第二概率小于所述第二概率阈值的情况下，判定所述报文为异常报文。

12.根据权利要求8所述的装置，其特征在于，还包括：

第二获取模块，用于在车辆正常运行过程中，获取所述车辆的CAN总线上的样本报文；

第二解析模块，用于解析所述样本报文，以获得样本数据序列；

训练模块，用于基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型。

13.根据权利要求12所述的装置，所述报文检测模型包括参数检测模型，所述训练模块包括：

第三确定子模块，用于从所述样本数据序列中确定多个样本字符类型序列，一个样本字符类型序列对应一个参数，所述样本字符类型序列包括对应参数中每一位的字符类型；

第一训练子模块，用于基于各样本字符类型序列分别训练隐式马尔可夫模型，以得到多个参数检测模型，一个参数检测模型对应一个参数。

14.根据权利要求12所述的装置，其特征在于，所述报文检测模型包括指令检测模型，所述训练模块包括：

第四确定子模块，用于从所述样本数据序列中确定样本指令序列，所述样本指令序列包括多条按时间排序的样本指令；

第二训练子模块，用于基于所述样本指令序列训练隐式马尔可夫模型，以得到指令检测模型。

15.一种报文检测设备，其特征在于，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或多个处理器实现如权利要求1至7中任一项所述的方法。

16.一种计算机可读存储介质，其存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求1至7中任一项所述的方法。

报文检测方法、装置、设备和存储介质

技术领域

[0001] 本发明涉及车辆安全领域，尤其涉及一种报文检测方法、装置、设备和存储介质。

背景技术

[0002] 控制器局域网络(Controller Area Network，CAN)总线协议已经成为车辆计算机控制系统和嵌入式工业控制局域网的标准总线。目前出现了一些通过入侵CAN总线而攻击车辆的行为，对车辆正常行驶构成了极大的安全威胁。目前常见的基于CAN总线的入侵检测都是基于规则和签名，即黑名单的形式。这种方式仅能针对已经出现的入侵模式进行防御，非常容易绕过和漏报。

发明内容

[0003] 本发明实施例提供一种报文检测方法、装置、设备和存储介质，以解决现有技术中的一个或多个技术问题。

[0004] 第一方面，本发明实施例提供了一种报文检测方法，包括：

[0005] 获取CAN总线上的报文；

[0006] 解析所述报文，以获得数据序列；

[0007] 根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文；

[0008] 其中，所述报文检测模型采用样本报文中的样本数据序列训练得到。

[0009] 在一种实施方式中，所述报文检测模型包括参数检测模型，根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文，包括：

[0010] 从所述数据序列中确定至少一个字符类型序列，所述字符类型序列包括目标参数中每一位的字符类型，所述目标参数为与所述字符类型序列对应的参数；

[0011] 根据所述目标参数对应的参数检测模型，判断所述报文是否为异常报文。

[0012] 在一种实施方式中，根据所述目标参数对应的参数检测模型，判断所述报文是否为异常报文，包括：

[0013] 将所述字符类型序列输入所述目标参数对应的参数检测模型；

[0014] 获取所述目标参数对应的参数检测模型输出的第一概率；

[0015] 判断所述第一概率是否小于所述目标参数对应的第一概率阈值；

[0016] 如果小于所述第一概率阈值，则判定所述报文为异常报文。

[0017] 在一种实施方式中，所述报文检测模型包括指令检测模型，根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文，包括：

[0018] 从所述数据序列中确定指令序列，所述指令序列包括多条按时间排序的指令；

[0019] 根据所述指令检测模型，判断所述报文是否为异常报文。

[0020] 在一种实施方式中，根据所述指令检测模型，判断所述报文是否为异常报文，包括：

[0021] 将所述指令序列输入所述指令检测模型；

[0022] 获取所述指令检测模型输出的第二概率；

[0023] 判断所述第二概率是否小于第二概率阈值；

[0024] 如果小于所述第二概率阈值，则判定所述报文为异常报文。

[0025] 在一种实施方式中，所述报文检测方法还包括：

[0026] 在车辆正常运行过程中，获取所述车辆的CAN总线上的样本报文；

[0027] 解析所述样本报文，以获得样本数据序列；

[0028] 基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型。

[0029] 在一种实施方式中，基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型，包括：

[0030] 从所述样本数据序列中确定多个样本字符类型序列，一个样本字符类型序列对应一个参数，所述样本字符类型序列包括对应参数中每一位的字符类型；

[0031] 基于各样本字符类型序列分别训练隐式马尔可夫模型，以得到多个参数检测模型，一个参数检测模型对应一个参数。

[0032] 在一种实施方式中，基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型，包括：

[0033] 从所述样本数据序列中确定样本指令序列，所述样本指令序列包括多条按时间排序的样本指令；

[0034] 基于所述样本指令序列训练隐式马尔可夫模型，以得到指令检测模型。

[0035] 第二方面，本发明实施例提供一种报文检测装置，包括：

[0036] 第一获取模块，用于获取CAN总线上的报文；

[0037] 第一解析模块，用于解析所述报文，以获得数据序列；

[0038] 判断模块，用于根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文；

[0039] 其中，所述报文检测模型采用样本报文中的样本数据序列训练得到。

[0040] 在一种实施方式中，所述报文检测模型包括参数检测模型，所述判断模块包括：

[0041] 第一确定子模块，用于从所述数据序列中确定至少一个字符类型序列，所述字符类型序列包括目标参数中每一位的字符类型，所述目标参数为与所述字符类型序列对应的参数；

[0042] 第一判断子模块，用于根据所述目标参数对应的参数检测模型，判断所述报文是否为异常报文。

[0043] 在一种实施方式中，所述第一判断子模块包括：

[0044] 第一输入单元，用于将所述字符类型序列输入所述目标参数对应的参数检测模型；

[0045] 第一获取单元，用于获取所述目标参数对应的参数检测模型输出的第一概率；

[0046] 第一判断单元，用于判断所述第一概率是否小于所述目标参数对应的第一概率阈值；

[0047] 第一判定单元，用于在所述第一概率小于所述第一概率阈值的情况下，判定所述报文为异常报文。

[0048] 在一种实施方式中，所述报文检测模型包括指令检测模型，所述判断模块包括：

[0049] 第二确定子模块，用于从所述数据序列中确定指令序列，所述指令序列包括多条按时间排序的指令；

[0050] 第二判断子模块，用于根据所述指令检测模型，判断所述报文是否为异常报文。

[0051] 在一种实施方式中，所述第二判断子模块包括：

[0052] 第二输入单元，用于将所述指令序列输入所述指令检测模型；

[0053] 第二获取单元，用于获取所述指令检测模型输出的第二概率；

[0054] 第二判断单元，用于判断所述第二概率是否小于第二概率阈值；

[0055] 第二判定单元，用于在所述第二概率小于所述第二概率阈值的情况下，判定所述报文为异常报文。

[0056] 在一种实施方式中，所述报文检测装置还包括：

[0057] 第二获取模块，用于在车辆正常运行过程中，获取所述车辆的CAN总线上的样本报文；

[0058] 第二解析模块，用于解析所述样本报文，以获得样本数据序列；

[0059] 训练模块，用于基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型。

[0060] 在一种实施方式中，所述报文检测模型包括参数检测模型，所述训练模块包括：

[0061] 第三确定子模块，用于从所述样本数据序列中确定多个样本字符类型序列，一个样本字符类型序列对应一个参数，所述样本字符类型序列包括对应参数中每一位的字符类型；

[0062] 第一训练子模块，用于基于各样本字符类型序列分别训练隐式马尔可夫模型，以得到多个参数检测模型，一个参数检测模型对应一个参数。

[0063] 在一种实施方式中，所述报文检测模型包括指令检测模型，所述训练模块包括：

[0064] 第四确定子模块，用于从所述样本数据序列中确定样本指令序列，所述样本指令序列包括多条按时间排序的样本指令；

[0065] 第二训练子模块，用于基于所述样本指令序列训练隐式马尔可夫模型，以得到指令检测模型。

[0066] 第三方面，本发明实施例提供了一种报文检测设备，所述设备的功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。所述硬件或软件包括一个或多个与上述功能相对应的模块。

[0067] 在一个可能的设计中，所述设备的结构中包括处理器和存储器，所述存储器用于存储支持所述装置执行上述方法的程序，所述处理器被配置为用于执行所述存储器中存储的程序。所述设备还可以包括通信接口，用于与其他设备或通信网络通信。

[0068] 第四方面，本发明实施例提供了一种计算机可读存储介质，用于存储报文检测装置所用的计算机软件指令，其包括用于执行上述方法所涉及的程序。

[0069] 上述技术方案通过监听车辆正常运行过程中产生的CAN总线报文，基于隐式马尔可夫模型训练出指令检测模型以及多个参数检测模型，可以识别出异常报文，进而发现对CAN总线的入侵行为。本发明实施例的方法不依赖于静态的规则，可以发现未知的入侵行为，避免绕过和漏报，还可以减少因人工维护规则而造成的成本浪费。

[0070] 上述概述仅仅是为了说明书的目的，并不意图以任何方式进行限制。除上述描述的示意性的方面、实施方式和特征之外，通过参考附图和以下的详细描述，本发明进一步的方面、实施方式和特征将会是容易明白的。

附图说明

[0071] 在附图中，除非另外规定，否则贯穿多个附图相同的附图标记表示相同或相似的部件或元素。这些附图不一定是按照比例绘制的。应该理解，这些附图仅描绘了根据本发明公开的一些实施方式，而不应将其视为是对本发明范围的限制。

[0072] 图1示出根据本发明实施例的报文检测方法的流程图。

[0073] 图2示出根据本发明实施例一种实施方式的报文检测方法的流程图。

[0074] 图3示出根据本发明实施例另一种实施方式的报文检测方法的流程图。

[0075] 图4示出根据本发明实施例又一种实施方式的报文检测方法的流程图。

[0076] 图5示出根据本发明实施例再一种实施方式的报文检测方法的流程图。

[0077] 图6示出根据本发明实施例再一种实施方式的报文检测方法的流程图。

[0078] 图7示出根据本发明实施例的报文检测装置的结构框图。

[0079] 图8示出根据本发明实施例一种实施方式的报文检测装置的结构框图。

[0080] 图9示出根据本发明实施例另一种实施方式的报文检测装置的结构框图。

[0081] 图10示出根据本发明实施例又一种实施方式的报文检测装置的结构框图。

[0082] 图11示出根据本发明实施例的报文检测设备的结构框图。

具体实施方式

[0083] 在下文中，仅简单地描述了某些示例性实施例。正如本领域技术人员可认识到的那样，在不脱离本发明的精神或范围的情况下，可通过各种不同方式修改所描述的实施例。因此，附图和描述被认为本质上是示例性的而非限制性的。

[0084] 图1示出根据本发明实施例的报文检测方法的流程图。如图1所示，该方法可以包括以下步骤：

[0085] 步骤S101、获取CAN总线上的报文；

[0086] 步骤S102、解析所述报文，以获得数据序列；

[0087] 步骤S103、根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文；

[0088] 其中，所述报文检测模型采用样本报文中的样本数据序列训练得到。

[0089] 通过CAN总线，可以控制车辆的发动机、变速箱、制动防抱死系统(Antilock Brake System，ABS)等车身安全模块，并将转速、车速、油温等数据共享至全车，实现车辆智能化控制。例如：高速行驶时自动锁闭车门；安全气囊弹出时自动开启车门。

[0090] CAN总线上的数据可以以报文的形式传输。报文包括但不限于数据帧、远程帧(总线发出)、错误帧和过载帧(用以提供附加的延时)。在一段时间内，监听CAN总线上的报文，并解析报文，可以获得多个数据序列。

[0091] 在一个示例中，报文检测模型可以有多个，需要根据所获得的数据序列，确定与该数据序列对应的报文检测模型，进而采用该报文检测模型对所获得的数据序列进行检测，

并根据检测结果判断报文是否为异常报文。如果有异常报文出现，则可以判定CAN总线被入侵。

[0092] 在一种实施方式中，如图2所示，在步骤S103中，可以包括：

[0093] 步骤S201、从所述数据序列中确定至少一个字符类型序列，所述字符类型序列包括目标参数中每一位的字符类型，所述目标参数为与所述字符类型序列对应的参数；

[0094] 步骤S202、根据所述目标参数对应的参数检测模型，判断所述报文是否为异常报文。

[0095] CAN总线的报文可以解析成指令和参数两部分。比如报文“加速到60公里/小时”可以解析为指令“加速”和参数“60”。数据序列可以包括字符类型序列。对于每个参数，可以根据其结构定义字符类型序列。例如：参数“60”有两位字符构成，第一位的字符“6”和第二位的字符“0”都为数字型。因此，参数“60”对应的字符类型序列即为“数字型，数字型”。

[0096] 本发明实施例中，字符类型可以包括数字型、字母型以及特殊符号型。其中，标点符号、单位符号、希腊字母或其他可用美国信息交换标准代码(American Standard Code for Information Interchange，ASCII)表示的字符可以归为特殊符号型。

[0097] 通常，参数的字符类型序列可以为标准形式。例如：参数中的每一位都是数字型(全数字型)；或者，每一位都是字母型(全字母型)；或者，第一位是数字型，第二位是数字型，第三位是字母型。如果一个参数的标准形式为全数字型的，突然出现了某一位是字母型的情况，那么说明该参数出现了异常，进而可以判断报文为异常报文。

[0098] 在一个示例中，数据序列可以包括多个字符类型序列，如S1、S2和S3。各字符类型序列分别对应一个参数。与某一字符类型序列对应的参数，可以称作该字符类型序列的目标参数。例如：与字符类型序列S1对应的参数F1，即为字符类型序列S1的目标参数。类似地，可以得到字符类型序列S2的目标参数F2，字符类型序列S3的目标参数F3。

[0099] 在一个示例中，报文检测模型可以包括多个参数检测模型。其中，每个参数分别对应一个参数检测模型。

[0100] 进一步地，可以根据目标参数F1，确定与目标参数F1对应的参数检测模型M1；然后采用参数检测模型M1检测字符类型序列S1，以判断目标参数F1的标准形式是否出现异常，进而判断报文是否为异常报文。或者，可以根据目标参数F2，确定与目标参数F2对应的参数检测模型M2；然后采用参数检测模型M2检测字符类型序列S2，以判断目标参数F2的标准形式是否出现异常，进而判断报文是否为异常报文。或者，可以根据目标参数F3，确定与目标参数F3对应的参数检测模型M3；然后采用参数检测模型M3检测字符类型序列S3，以判断目标参数F3的标准形式是否出现异常，进而判断报文是否为异常报文。

[0101] 在一种实施方式中，如图3所示，在步骤S202中可以包括：

[0102] 步骤S301、将所述字符类型序列输入所述目标参数对应的参数检测模型；

[0103] 步骤S302、获取所述目标参数对应的参数检测模型输出的第一概率；

[0104] 步骤S303、判断所述第一概率是否小于所述目标参数对应的第一概率阈值；如果小于所述第一概率阈值，则进入步骤S304；

[0105] 步骤S304、判定所述报文为异常报文。

[0106] 在一个示例中，目标参数F1可以对应第一概率阈值T1；目标参数F2可以对应第一概率阈值T2；目标参数F3可以对应第一概率阈值T3。

[0107] 可以将字符类型序列S1输入参数检测模型M1，得到第一概率P1；如果P1小于T1，则可以判断报文为异常报文。或者，可以将字符类型序列S2输入参数检测模型M2，得到第一概率P2；如果P2小于T2，则可以判断报文为异常报文。或者，可以将字符类型序列S3输入参数检测模型M3，得到第一概率P3；如果P3小于T3，则可以判断报文为异常报文。

[0108] 在一种实施方式中，报文检测模型可以包括指令检测模型。如图4所示，在步骤S103中可以包括：

[0109] 步骤S401、从所述数据序列中确定指令序列，所述指令序列包括多条按时间排序的指令；

[0110] 步骤S402、根据所述指令检测模型，判断所述报文是否为异常报文。

[0111] CAN总线的报文可以解析成指令和参数两部分。比如报文“加速到60公里/小时”可以解析为指令“加速”和参数“60”。数据序列可以包括指令序列。其中，指令序列可以包括多条指令，多条指令可以按照各指令的时间信息排列。指令的时间信息可以设定为该指令的获得时间，也可以设定为该指令在CAN总线上的发送时间，本发明实施例不作限定，只要在设定时可以统一即可。

[0112] 采用指令检测模型可以检测指令序列中各指令的内容或顺序是否异常，以判断报文是否为异常报文。

[0113] 在一个示例中，指令检测模型可以检测的指令序列的最大长度为5(5个指令)。如果获得的指令序列超过最大长度，例如为10，那么可以将该指令序列拆分成两个长度为5的指令序列。然后，采用指令检测模型分别检测这两个长度为5的指令序列。

[0114] 在一种实施方式中，如图5所示，在步骤S402中可以包括：

[0115] 步骤S501、将所述指令序列输入所述指令检测模型；

[0116] 步骤S502、获取所述指令检测模型输出的第二概率；

[0117] 步骤S503、判断所述第二概率是否小于第二概率阈值；如果小于所述第二概率阈值，则进入步骤S504；

[0118] 步骤S504、判定所述报文为异常报文。

[0119] 指令检测模型可以对应有第二概率阈值。如果第一概率小于第二概率阈值，则可以判断指令序列出现异常，进而判断报文为异常报文。

[0120] 本发明实施例中，报文检测模型可以采用样本报文中的样本数据序列训练得到。下面介绍报文检测模型的训练方法。

[0121] 在一种实施方式中，如图6所示，本发明实施例的方法还可以包括：

[0122] 步骤S601、在车辆正常运行过程中，获取所述车辆的CAN总线上的样本报文；

[0123] 步骤S602、解析所述样本报文，以获得样本数据序列；

[0124] 步骤S603、基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型。

[0125] 隐式马尔可夫模型(Hidden Markov Model，HMM)是统计模型，它用来描述含有隐含未知参数的马尔可夫过程。在隐式马尔可夫模型中，每个观测向量都是通过某些概率密度分布表现为各种状态，每一个观测向量是由一个具有相应概率密度分布的状态序列产生。

[0126] 长期监听车辆正常运行过程中CAN总线上的报文，并解析报文，以获得大量样本数

据序列。用这些样本数据序列训练隐式马尔可夫模型，可以获得报文检测模型。报文检测模型可以表征正常报文规则和报文白名单。

[0127] 在一种实施方式中，样本数据序列可以包括多个样本字符类型序列。一个样本字符类型序列对应一个参数。样本字符类型序列包括与该样本字符类型序列对应的参数中每一位的字符类型。基于各样本字符类型序列分别训练隐式马尔可夫模型，以得到多个参数检测模型，一个参数检测模型对应一个参数。

[0128] 在一个示例中，对于参数F1，可以用大量与参数F1对应的样本字符类型序列来训练隐式马尔可夫模型，以得到与参数F1对应的参数检测模型M1。对于参数F2，可以用大量与参数F2对应的样本字符类型序列来训练隐式马尔可夫模型，以得到与参数F2对应的参数检测模型M2。对于参数F3，可以用大量与参数F3对应的样本字符类型序列来训练隐式马尔可夫模型，以得到与参数F3对应的参数检测模型M3。

[0129] 在一种实施方式中，样本数据序列可以包括样本指令序列。样本指令序列可以包括多条按时间排序的样本指令。从样本数据序列中确定样本指令序列，进而用样本指令序列训练隐式马尔可夫模型，可以得到指令检测模型。

[0130] 综上所述，本实施例的报文检测方法，通过监听车辆正常运行过程中产生的CAN总线报文，基于隐式马尔可夫模型训练出指令检测模型以及多个参数检测模型，可以识别出异常报文，如指令级别和参数级别的异常报文，进而发现对CAN总线的入侵行为。本发明实施例的方法不用依赖于静态的规则，可以发现未知的入侵行为，避免绕过和漏报，还可以减少因人工维护规则而造成的成本浪费。

[0131] 图7示出根据本发明实施例的报文检测装置的结构框图。如图7所示，该装置可以包括：

[0132] 第一获取模块701，用于获取CAN总线上的报文；

[0133] 第一解析模块702，用于解析所述报文，以获得数据序列；

[0134] 判断模块703，用于根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文；

[0135] 其中，所述报文检测模型采用样本报文中的样本数据序列训练得到。

[0136] 在一种实施方式中，如图8所示，所述报文检测模型可以包括参数检测模型，判断模块703可以包括：

[0137] 第一确定子模块801，用于从所述数据序列中确定至少一个字符类型序列，所述字符类型序列包括目标参数中每一位的字符类型，所述目标参数为与所述字符类型序列对应的参数；

[0138] 第一判断子模块802，用于根据所述目标参数对应的参数检测模型，判断所述报文是否为异常报文。

[0139] 在一种实施方式中，如图8所示，第一判断子模块802可以包括：

[0140] 第一输入单元821，用于将所述字符类型序列输入所述目标参数对应的参数检测模型；

[0141] 第一获取单元822，用于获取所述目标参数对应的参数检测模型输出的第一概率；

[0142] 第一判断单元823，用于判断所述第一概率是否小于所述目标参数对应的第一概率阈值；

[0143] 第一判定单元824，用于在所述第一概率小于所述第一概率阈值的情况下，判定所述报文为异常报文。

[0144] 在一种实施方式中，如图9所示，所述报文检测模型可以包括指令检测模型，判断模块703可以包括：

[0145] 第二确定子模块901，用于从所述数据序列中确定指令序列，所述指令序列包括多条按时间排序的指令；

[0146] 第二判断子模块902，用于根据所述指令检测模型，判断所述报文是否为异常报文。

[0147] 在一种实施方式中，如图9所示，第二判断子模块902可以包括：

[0148] 第二输入单元921，用于将所述指令序列输入所述指令检测模型；

[0149] 第二获取单元922，用于获取所述指令检测模型输出的第二概率；

[0150] 第二判断单元923，用于判断所述第二概率是否小于第二概率阈值；

[0151] 第二判定单元924，用于在所述第二概率小于所述第二概率阈值的情况下，判定所述报文为异常报文。

[0152] 在一种实施方式中，如图10所示，本发明实施例的报文检测装置还可以包括：

[0153] 第二获取模块1001，用于在车辆正常运行过程中，获取所述车辆的CAN总线上的样本报文；

[0154] 第二解析模块1002，用于解析所述样本报文，以获得样本数据序列；

[0155] 训练模块1003，用于基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型。

[0156] 在一种实施方式中，如图10所示，所述报文检测模型包括参数检测模型，训练模块1003可以包括：

[0157] 第三确定子模块1031，用于从所述样本数据序列中确定多个样本字符类型序列，一个样本字符类型序列对应一个参数，所述样本字符类型序列包括对应参数中每一位的字符类型；

[0158] 第一训练子模块1032，用于基于各样本字符类型序列分别训练隐式马尔可夫模型，以得到多个参数检测模型，一个参数检测模型对应一个参数。

[0159] 在一种实施方式中，如图10所示，所述报文检测模型包括指令检测模型，训练模块1003可以包括：

[0160] 第四确定子模块1033，用于从所述样本数据序列中确定样本指令序列，所述样本指令序列包括多条按时间排序的样本指令；

[0161] 第二训练子模块1034，用于基于所述样本指令序列训练隐式马尔可夫模型，以得到指令检测模型。

[0162] 本发明实施例各装置中的各模块的功能可以参见上述方法中的对应描述，在此不再赘述。

[0163] 图11示出根据本发明实施例的报文检测设备的结构框图。如图11所示，该设备包括：存储器1101和处理器1102，存储器1101内存储有可在处理器1102上执行的计算机程序。所述处理器1102执行所述计算机程序时实现上述实施例中的报文检测方法。所述存储器1101和处理器1102的数量可以为一个或多个。

[0164] 该设备还包括：

[0165] 通信接口1103，用于与外界设备进行通信，进行数据交互传输。

[0166] 存储器1101可能包含高速RAM存储器，也可能还包括非易失性存储器(non-volatile memory)，例如至少一个磁盘存储器。

[0167] 如果存储器1101、处理器1102和通信接口1103独立实现，则存储器1101、处理器1102和通信接口1103可以通过总线相互连接并完成相互间的通信。所述总线可以是工业标准体系结构(ISA，Industry Standard Architecture)总线、外部设备互连(PCI，Peripheral Component Interconnect)总线或扩展工业标准体系结构(EISA，Extended Industry Standard Component)总线等。所述总线可以分为地址总线、数据总线、控制总线等。为便于表示，图11中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。

[0168] 可选的，在具体实现上，如果存储器1101、处理器1102及通信接口1103集成在一块芯片上，则存储器1101、处理器1102及通信接口1103可以通过内部接口完成相互间的通信。

[0169] 本发明实施例提供了一种计算机可读存储介质，其存储有计算机程序，该程序被处理器执行时实现上述实施例中任一所述的方法。

[0170] 在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0171] 此外，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或隐含地包括至少一个该特征。在本发明的描述中，“多个”的含义是两个或两个以上，除非另有明确具体的限定。

[0172] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本发明的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本发明的实施例所属技术领域的技术人员所理解。

[0173] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤，例如，可以被认为是用于实现逻辑功能的可执行指令的定序列表，可以具体实现在任何计算机可读介质中，以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用，或结合这些指令执行系统、装置或设备而使用。就本说明书而言，“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例(非穷尽性列表)包括以下：具有一个或多个布线的电连接部(电子装置)，便携式计算机盘盒(磁装置)，随机存取存储器(RAM)，只读存储器(ROM)，可擦除可编辑只读存储器(EPROM或闪速存储器)，光纤装置，以及便携式只读存储器(CDROM)。另外，计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介

质，因为可以例如通过对纸或其他介质进行光学扫描，接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序，然后将其存储在计算机存储器中。

[0174] 应当理解，本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列(PGA)，现场可编程门阵列(FPGA)等。

[0175] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

[0176] 此外，在本发明各个实施例中的各功能单元可以集成在一个处理模块中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，也可以存储在一个计算机可读存储介质中。所述存储介质可以是只读存储器，磁盘或光盘等。

[0177] 以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到其各种变化或替换，这些都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以所述权利要求的保护范围为准。

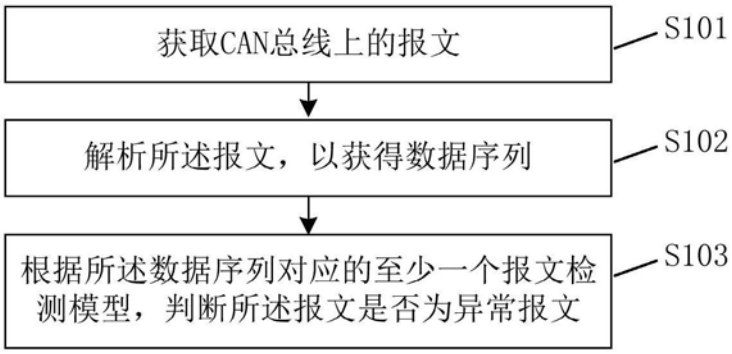

获取CAN总线上的报文
S101
解析所述报文，以获得数据序列
S102
根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文
S103

图1

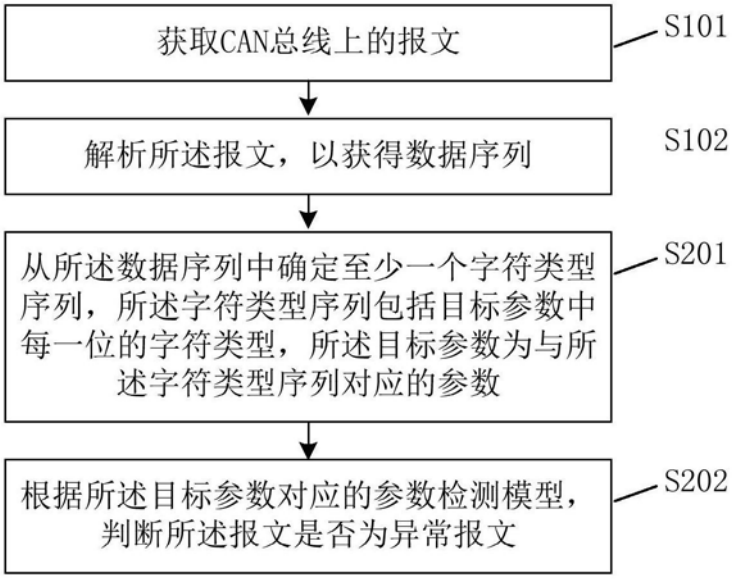

获取CAN总线上的报文
S101
解析所述报文，以获得数据序列
S102
从所述数据序列中确定至少一个字符类型序列，所述字符类型序列包括目标参数中每一位的字符类型，所述目标参数为与所述字符类型序列对应的参数
S201
根据所述目标参数对应的参数检测模型，判断所述报文是否为异常报文
S202

图2

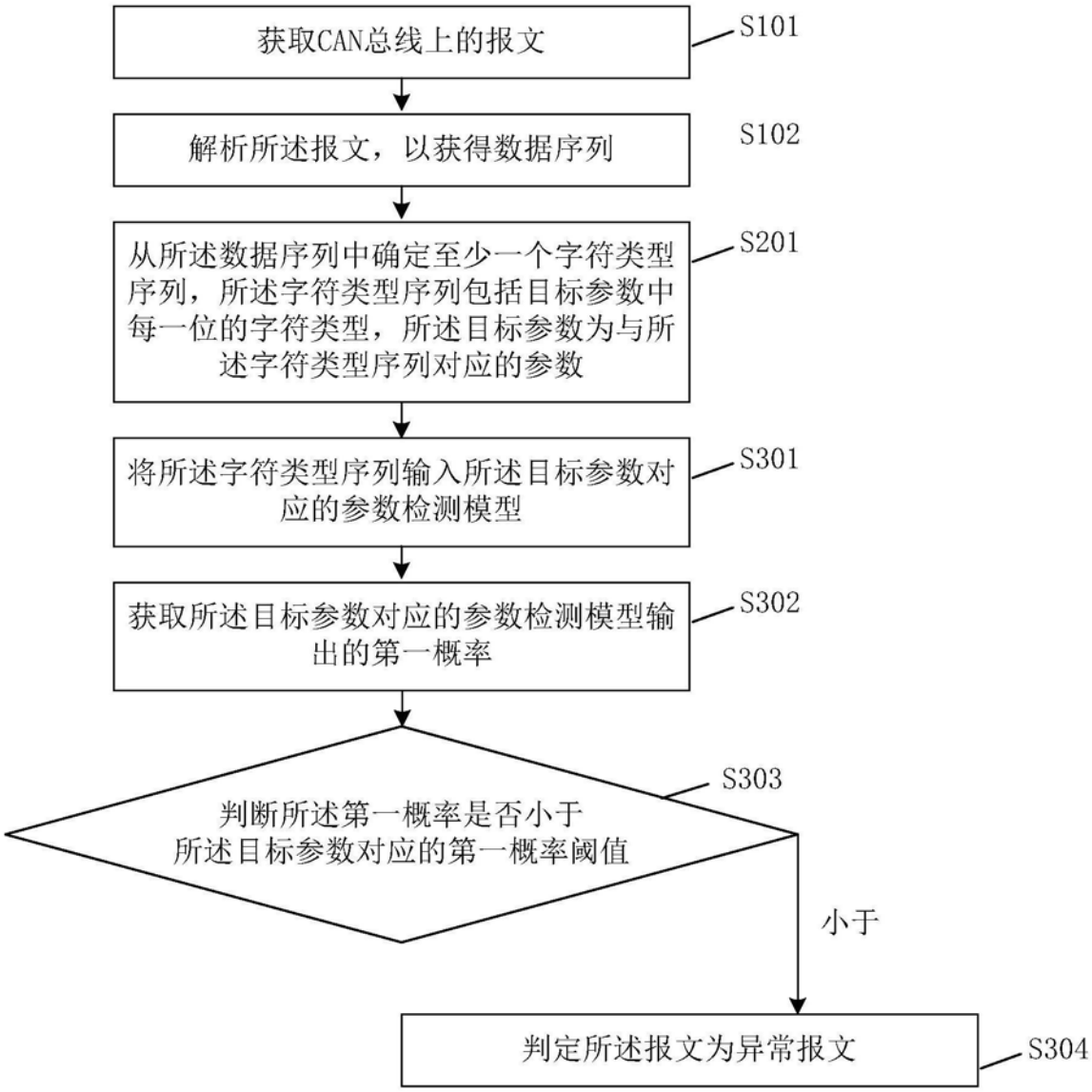

获取CAN总线上的报文
S101
解析所述报文，以获得数据序列
S102
从所述数据序列中确定至少一个字符类型序列，所述字符类型序列包括目标参数中每一位的字符类型，所述目标参数为与所述字符类型序列对应的参数
S201
将所述字符类型序列输入所述目标参数对应的参数检测模型
S301
获取所述目标参数对应的参数检测模型输出的第一概率
S302
判断所述第一概率是否小于所述目标参数对应的第一概率阈值
S303
小于
判定所述报文为异常报文
S304

图3

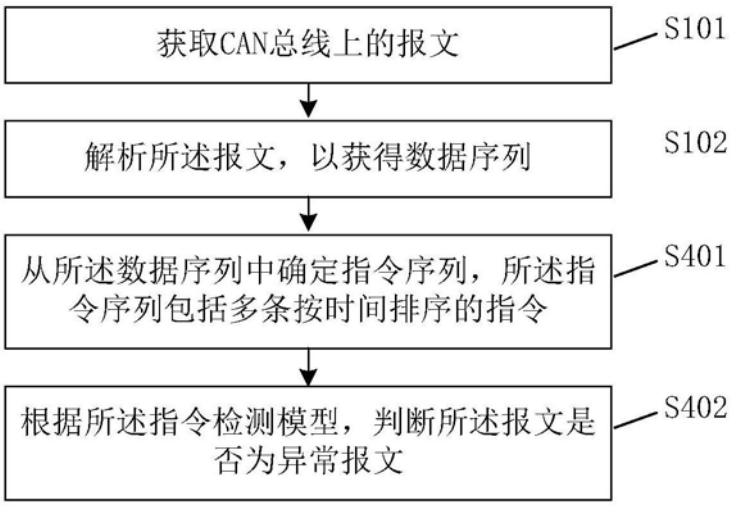

获取CAN总线上的报文
S101
解析所述报文，以获得数据序列
S102
从所述数据序列中确定指令序列，所述指令序列包括多条按时间排序的指令
S401
根据所述指令检测模型，判断所述报文是否为异常报文
S402

图4

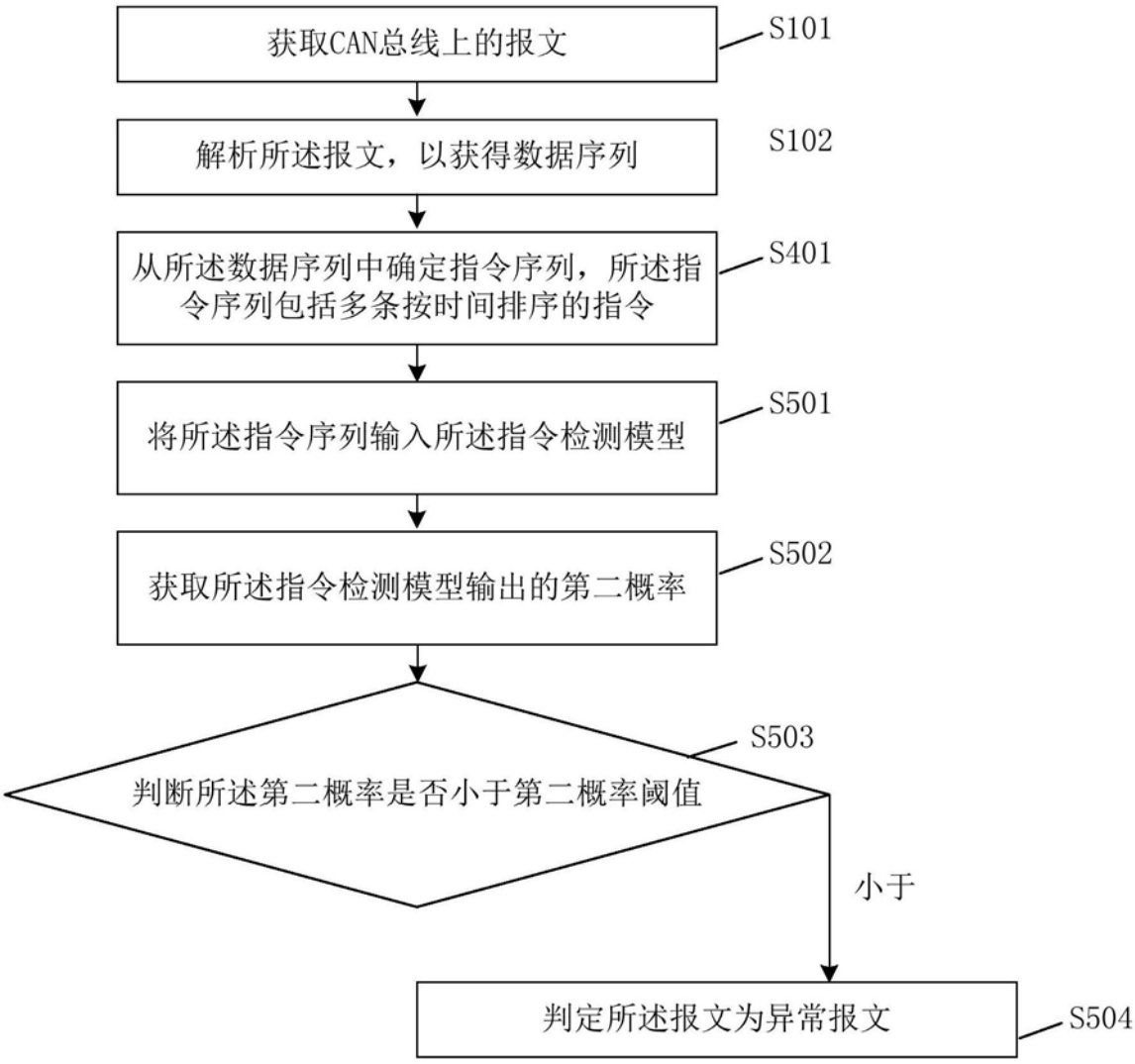

获取CAN总线上的报文
S101
解析所述报文，以获得数据序列
S102
从所述数据序列中确定指令序列，所述指令序列包括多条按时间排序的指令
S401
将所述指令序列输入所述指令检测模型
S501
获取所述指令检测模型输出的第二概率
S502
判断所述第二概率是否小于第二概率阈值
S503
小于
判定所述报文为异常报文
S504

图5

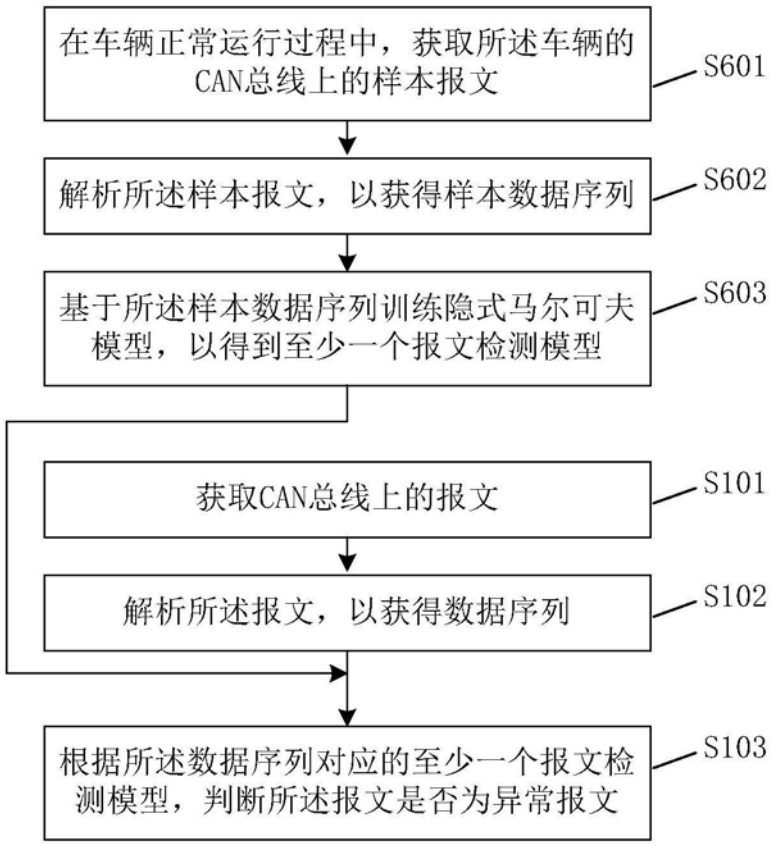

在车辆正常运行过程中，获取所述车辆的CAN总线上的样本报文
S601
解析所述样本报文，以获得样本数据序列
S602
基于所述样本数据序列训练隐式马尔可夫模型，以得到至少一个报文检测模型
S603
获取CAN总线上的报文
S101
解析所述报文，以获得数据序列
S102
根据所述数据序列对应的至少一个报文检测模型，判断所述报文是否为异常报文
S103

图6


第一获取模块
701
第一解析模块
702
判断模块
703

图7

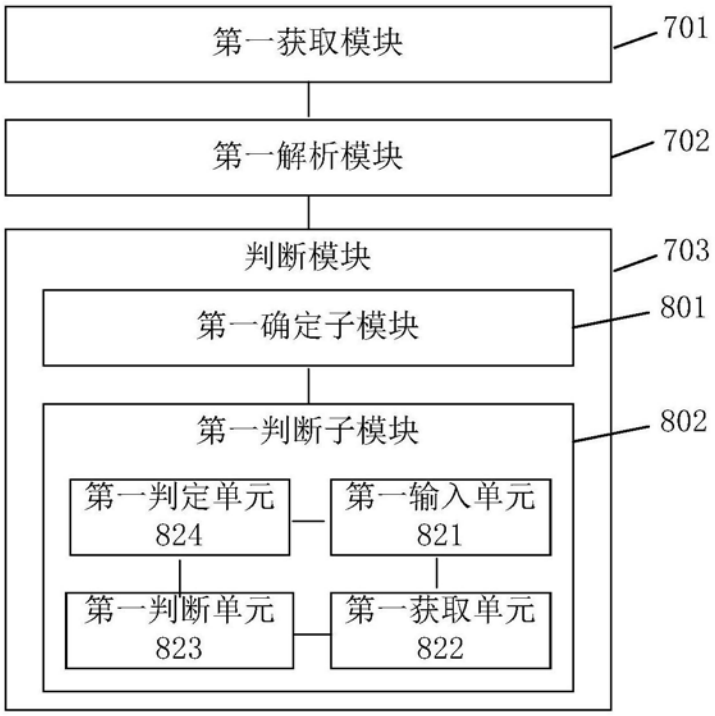

第一获取模块
701
第一解析模块
702
判断模块
703
第一确定子模块
801
第一判断子模块
802
第一判定单元
824
第一输入单元
821
第一判断单元
823
第一获取单元
822

图8

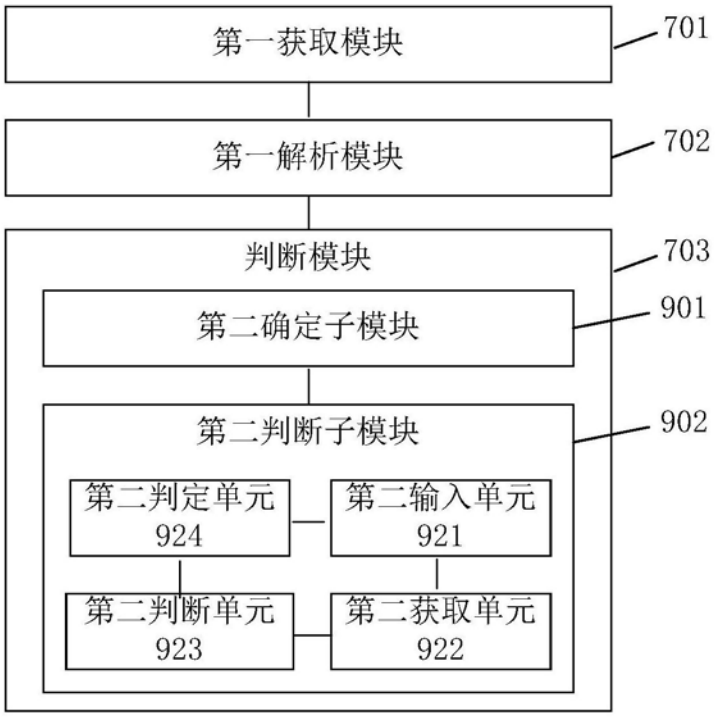

第一获取模块
701
第一解析模块
702
判断模块
703
第二确定子模块
901
第二判断子模块
902
第二判定单元
924
第二输入单元
921
第二判断单元
923
第二获取单元
922

图9

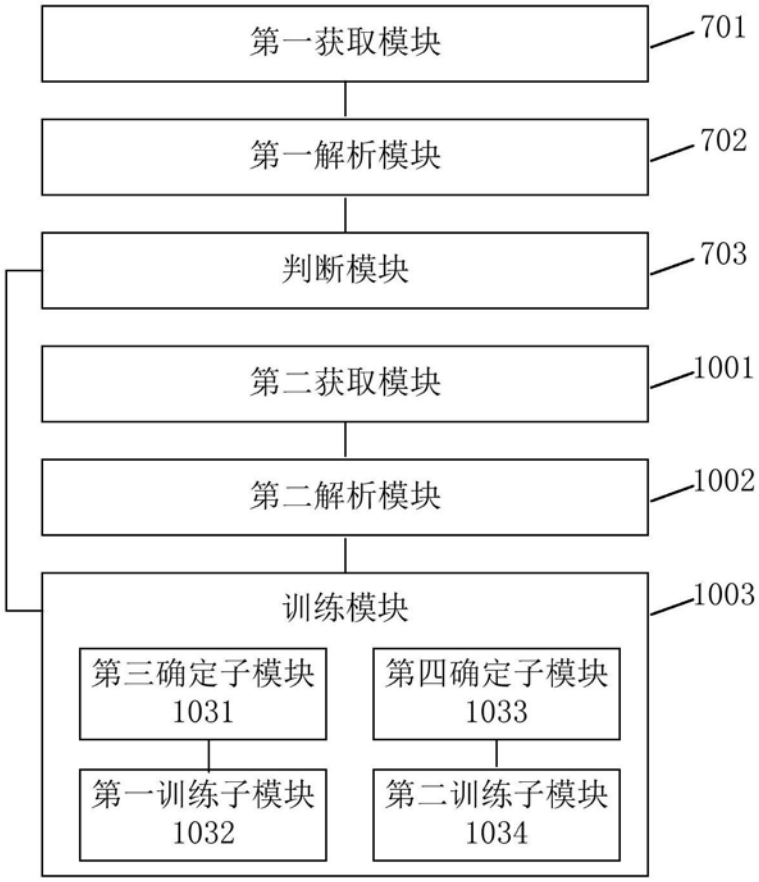

第一获取模块
701
第一解析模块
702
判断模块
703
第二获取模块
1001
第二解析模块
1002
训练模块
1003
第三确定子模块
1031
第四确定子模块
1033
第一训练子模块
1032
第二训练子模块
1034

图10

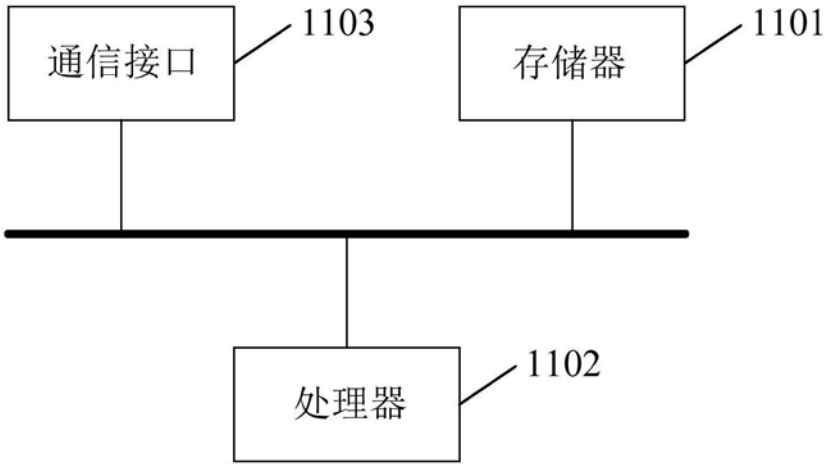

通信接口
1103
存储器
1101
处理器
1102

图11