(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116112418 A
(43)申请公布日 2023.05.12

(21)申请号 202310079065.9
(22)申请日 2023.01.17
(71)申请人 清华大学
地址 100084 北京市海淀区清华园1号
(72)发明人 李江 曹家浩 孟子立 谢仁杰 徐明伟
(74)专利代理机构 北京清亦华知识产权代理事务所(普通合伙) 11201
专利代理师 黄德海
(51)Int.Cl.
H04L 45/00(2022.01)
H04L 45/02(2022.01)
H04L 45/021(2022.01)
H04L 45/28(2022.01)
H04L 45/74(2022.01)
H04L 45/745(2022.01)
H04L 9/40(2022.01)

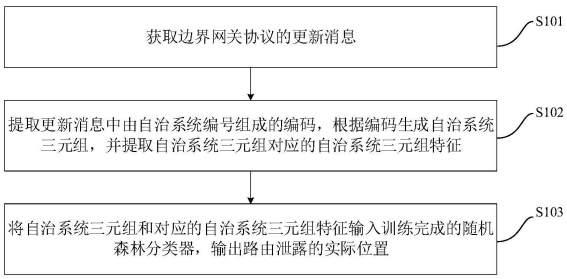
权利要求书2页 说明书10页 附图2页

(54)发明名称

路由泄露的定位方法、装置、电子设备及存储介质

(57)摘要

本申请涉及路由异常定位技术领域，特别涉及一种路由泄露的定位方法、装置、电子设备及存储介质，其中，方法包括：获取边界网关协议的更新消息；提取更新消息中由自治系统编号组成的编码，根据编码生成自治系统三元组，并提取自治系统三元组对应的自治系统三元组特征；将自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器，输出路由泄露的实际位置。由此，解决了相关技术中无法准确定位路由泄露位置，定位花费时间长等问题。



CN 116112418 A

1.一种路由泄露的定位方法,其特征在于,包括以下步骤:

获取边界网关协议的更新消息;

提取所述更新消息中由自治系统编号组成的编码,根据所述编码生成自治系统三元组,并提取所述自治系统三元组对应的自治系统三元组特征;

将所述自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器,输出路由泄露的实际位置。

2.根据权利要求1所述的方法,其特征在于,所述随机森林分类器的训练包括:

获取路由泄露事件,其中,所述路由泄露事件包括发生路由泄露的自治系统三元组和未发生路由泄露的自治系统三元组;

根据所述发生路由泄露的自治系统三元组、所述未发生路由泄露的自治系统三元组以及各自对应的自治系统三元组特征生成正负样本;

利用所述正负样本对随机森林分类器进行训练,直到满足训练停止条件,得到训练完成的随机森林分类器。

3.根据权利要求1所述的方法,其特征在于,所述提取所述自治系统三元组对应的自治系统三元组特征,包括:

以所述自治系统三元组为索引,查询预先建立的数据库,输出所述自治系统三元组特征。

4.根据权利要求3所述的方法,其特征在于,所述数据库存储有间隔预设时长更新的自治系统特征,其中,通过第一路由数据源更新所述自治系统特征中的自治系统距离、自治系统度和自治系统类型,通过第二路由数据源和第三路由数据源更新所述自治系统特征中的自治系统地址空间,通过所述第一路由数据源和第四路由数据源更新所述自治系统特征中的自治系统地理位置。

5.根据权利要求1所述的方法,其特征在于,所述根据所述编码生成自治系统三元组,包括:

识别所述编码中预设编码、重复编号和目标字段;

丢弃或移除所述预设编码,删除所述重复编码,对所述目标字段中包含的自治系统编号对应的自治系统生成各自对应的自治系统三元组,其中,对于满足预设条件的预设编码,丢弃所述编码对应的自治系统三元组。

6.根据权利要求1-5任意一项所述的方法,其特征在于,在输出路由泄露的实际位置之后,还包括:

生成路由泄露的报警信息,发送所述报警信息至预设终端。

7.一种路由泄露的定位装置,其特征在于,包括:

获取模块,用于获取边界网关协议的更新消息;

提取模块,用于提取所述更新消息中由自治系统编号组成的编码,根据所述编码生成自治系统三元组,并提取所述自治系统三元组对应的自治系统三元组特征;

输出模块,用于将所述自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器,输出路由泄露的实际位置。

8.根据权利要求7所述的装置,其特征在于,所述输出模块进一步用于:

获取路由泄露事件,其中,所述路由泄露事件包括发生路由泄露的自治系统三元组和

未发生路由泄露的自治系统三元组；

根据所述发生路由泄露的自治系统三元组、所述未发生路由泄露的自治系统三元组以及各自对应的自治系统三元组特征生成正负样本；

利用所述正负样本对随机森林分类器进行训练，直到满足训练停止条件，得到训练完成的随机森林分类器。

9.根据权利要求7所述的装置，其特征在于，所述提取模块进一步用于：

以所述自治系统三元组为索引，查询预先建立的数据库，输出所述自治系统三元组特征。

10.根据权利要求9所述的装置，其特征在于，所述数据库存储有间隔预设时长更新的自治系统特征，其中，通过第一路由数据源更新所述自治系统特征中的自治系统距离、自治系统度和自治系统类型，通过第二路由数据源和第三路由数据源更新所述自治系统特征中的自治系统地址空间，通过所述第一路由数据源和第四路由数据源更新所述自治系统特征中的自治系统地理位置。

11.根据权利要求7所述的装置，其特征在于，所述提取模块进一步用于：

识别所述编码中预设编码、重复编号和目标字段；

丢弃或移除所述预设编码，删除所述重复编码，对所述目标字段中包含的自治系统编号对应的自治系统生成各自对应的自治系统三元组，其中，对于满足预设条件的预设编码，丢弃所述编码对应的自治系统三元组。

12.根据权利要求7-11任意一项所述的装置，其特征在于，还包括：

发送模块，用于在输出路由泄露的实际位置之后，生成路由泄露的报警信息，发送所述报警信息至预设终端。

13.一种电子设备，其特征在于，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述程序，以实现如权利要求1-6任一项所述的路由泄露的定位方法。

14.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行，以用于实现如权利要求1-6任一项所述的路由泄露的定位方法。

路由泄露的定位方法、装置、电子设备及存储介质

技术领域

[0001] 本申请涉及路由异常定位技术领域，特别涉及一种路由泄露的定位方法、装置、电子设备及存储介质。

背景技术

[0002] BGP(Border Gateway Protocol，边界网关协议)是目前唯一真正使用的域间路由协议，其将全世界的各种网络，即自治系统(AS，Autonomous System)连接起来。各自治系统通过交换BGP更新消息学习到各目的地网络的可达性。BGP更新消息中含有去往目的地网络的AS_PATH，收到更新消息的自治系统可以沿着该消息中的AS_PATH去往目的地网络。AS_PATH可以理解为由一组自治系统的ASN(Autonomous System Number，自治系统编号)组成的编码。如果自治系统没有遵守正确的路由策略进行路由宣告，而是将路由宣告给了错误的自治系统，路由泄露就有可能发生。路由泄露可能会导致流量被攻击者监听或丢弃，大规模的路由泄露往往会造成全球范围的长时间网络性能下降甚至大型网络应用的瘫痪。

[0003] 互联网对于人类社会的重要性不言而喻。近些年国际互联网上发生了很多重大路由泄露事故，相应网络安全也成为重中之重。路由泄露的准确实时定位方法可以帮助网络管理员及时发现路由泄露位置，定位故障自治系统，生成相应的路由过滤器，最大化减少路由泄露事故带来的不良影响。

[0004] 研究人员提出了基于自治系统商业关系的路由泄漏实时定位方法。这类方法检查AS_PATH上的自治系统商业关系是否违背了无谷原则，即提取AS_PATH中的自治系统三元组，逐个检查每个自治系统三元组是否违背了无谷原则。这类方法尽管设计了可以较准确地推断出一对自治系统之间商业关系的算法，但是其无法准确定位路由泄露。定位路由泄露需要同时获知两对自治系统之间商业关系，根据概率乘法原则，这种方法同时准确获知两对自治系统之间商业关系的准确率将大大下降。此外，这类方法无法定位发生在其预先不可见的自治系统边上的路由泄露。

[0005] 另一类方法是将机器学习直接应用于大量BGP更新报文的统计特征，以此来检测路由泄漏。这类方法虽然实现了准确的路由泄漏检测，但是它们无法定位路由泄漏的位置，即发生泄露的自治系统，接收泄露路由的自治系统和泄露路由来源自治系统构成的三元组。因此，BGP安全专家仍然需要花费很长时间来定位路由泄漏。此外，这类方法需要用很长的时间来定期收集和计算统计特征，这将导致较大的检测延迟。综上，目前已有工作难以同时实现准确和实时的路由泄露定位。

发明内容

[0006] 本申请提供一种路由泄露的定位方法、装置、电子设备及存储介质，以解决相关技术中无法准确定位路由泄露位置，定位花费时间长等问题。

[0007] 本申请第一方面实施例提供一种路由泄露的定位方法，包括以下步骤：获取边界网关协议的更新消息；提取所述更新消息中由自治系统编号组成的编码，根据所述编码生

成自治系统三元组，并提取所述自治系统三元组对应的自治系统三元组特征；将所述自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器，输出路由泄露的实际位置。

[0008] 可选地，所述随机森林分类器的训练包括：获取路由泄露事件，其中，所述路由泄露事件包括发生路由泄露的自治系统三元组和未发生路由泄露的自治系统三元组；根据所述发生路由泄露的自治系统三元组、所述未发生路由泄露的自治系统三元组以及各自对应的自治系统三元组特征生成正负样本；利用所述正负样本对随机森林分类器进行训练，直到满足训练停止条件，得到训练完成的随机森林分类器。

[0009] 可选地，所述提取所述自治系统三元组对应的自治系统三元组特征，包括：以所述自治系统三元组为索引，查询预先建立的数据库，输出所述自治系统三元组特征。

[0010] 可选地，所述数据库存储有间隔预设时长更新的自治系统特征，其中，通过第一路由数据源更新所述自治系统特征中的自治系统距离、自治系统度和自治系统类型，通过第二路由数据源和第三路由数据源更新所述自治系统特征中的自治系统地址空间，通过所述第一路由数据源和第四路由数据源更新所述自治系统特征中的自治系统地理位置。

[0011] 可选地，所述根据所述编码生成自治系统三元组，包括：识别所述编码中预设编码、重复编号和目标字段；丢弃或移除所述预设编码，删除所述重复编码，对所述目标字段中包含的自治系统编号对应的自治系统生成各自对应的自治系统三元组，其中，对于满足预设条件的预设编码，丢弃所述编码对应的自治系统三元组。

[0012] 可选地，在输出路由泄露的实际位置之后，还包括：生成路由泄露的报警信息，发送所述报警信息至预设终端。

[0013] 本申请第二方面实施例提供一种路由泄露的定位装置，包括：获取模块，用于获取边界网关协议的更新消息；提取模块，用于提取所述更新消息中由自治系统编号组成的编码，根据所述编码生成自治系统三元组，并提取所述自治系统三元组对应的自治系统三元组特征；输出模块，用于将所述自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器，输出路由泄露的实际位置。

[0014] 可选地，所述输出模块进一步用于：获取路由泄露事件，其中，所述路由泄露事件包括发生路由泄露的自治系统三元组和未发生路由泄露的自治系统三元组；根据所述发生路由泄露的自治系统三元组、所述未发生路由泄露的自治系统三元组以及各自对应的自治系统三元组特征生成正负样本；利用所述正负样本对随机森林分类器进行训练，直到满足训练停止条件，得到训练完成的随机森林分类器。

[0015] 可选地，所述提取模块进一步用于：以所述自治系统三元组为索引，查询预先建立的数据库，输出所述自治系统三元组特征。

[0016] 可选地，所述数据库存储有间隔预设时长更新的自治系统特征，其中，通过第一路由数据源更新所述自治系统特征中的自治系统距离、自治系统度和自治系统类型，通过第二路由数据源和第三路由数据源更新所述自治系统特征中的自治系统地址空间，通过所述第一路由数据源和第四路由数据源更新所述自治系统特征中的自治系统地理位置。

[0017] 可选地，所述提取模块进一步用于：识别所述编码中预设编码、重复编号和目标字段；丢弃或移除所述预设编码，删除所述重复编码，对所述目标字段中包含的自治系统编号对应的自治系统生成各自对应的自治系统三元组，其中，对于满足预设条件的预设编码，丢

弃所述编码对应的自治系统三元组。

[0018] 可选地，还包括：发送模块，用于在输出路由泄露的实际位置之后，生成路由泄露的报警信息，发送所述报警信息至预设终端。

[0019] 本申请第三方面实施例提供一种电子设备，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述程序，以实现如上述实施例所述的路由泄露的定位方法。

[0020] 本申请第四方面实施例提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行，以用于实现如上述实施例所述的路由泄露的定位方法。

[0021] 由此，本申请至少具有如下有益效果：

[0022] 本申请实施例可以通过获得自治系统三元组特征以及训练随机森林分类器从BGP更新消息中的AS_PATH定位路由泄露；从单条BGP更新消息中提取AS_PATH，生成多个自治系统三元组，然后对逐个自治系统三元组判断其是否发生路由泄露；自治系统三元组特征能够准确区分路由泄露和正常路由；使用相对稳定，不需要频繁收集的自治系统三元组特征，节省定位时间，并基于轻量的随机森林分类器，能够实时地定位路由泄露位置。由此，解决了相关技术中无法准确定位路由泄露位置，定位花费时间长等技术问题。

[0023] 本申请附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本申请的实践了解到。

附图说明

[0024] 本申请上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

[0025] 图1为根据本申请实施例提供的路由泄露的定位方法的流程图；

[0026] 图2为根据本申请实施例提供的路由泄露的定位方法的示意图；

[0027] 图3为根据本申请实施例提供的定位泄露的延迟累积分布的示意图；

[0028] 图4为根据本申请实施例的路由泄露的定位装置的示意图；

[0029] 图5为根据本申请实施例提供的电子设备的结构示意图。

具体实施方式

[0030] 下面详细描述本申请的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，旨在用于解释本申请，而不能理解为对本申请的限制。

[0031] 下面参考附图描述本申请实施例的路由泄露的定位方法、装置、电子设备及存储介质。针对上述背景技术中提到的目前关于路由泄露的定位方法，无法准确定位路由泄露，需要很长时间来定期收集和计算统计特征的问题，本申请提供了一种路由泄露的定位方法，在该方法中，通过获得自治系统三元组特征以及训练随机森林分类器定位路由泄露，能够准确区分路由泄露和正常路由，使用相对稳定且不需要频繁收集的自治系统三元组特征。由此，解决了相关技术中无法准确定位路由泄露位置，定位花费时间长等问题。

[0032] 具体而言，图1为本申请实施例所提供的一种路由泄露的定位方法的流程示意图。

[0033] 如图1所示，该路由泄露的定位方法包括以下步骤：

[0034] 在步骤S101中,获取边界网关协议的更新消息。

[0035] 其中,边界网关协议(BGP,Border Gateway Protocol)是一种实现自治系统AS(Autonomous System)之间可达性信息交换的路径矢量路由协议,该协议主要通过更新消息传播路由信息。

[0036] 自治系统(AS,Autonomous System):互联网是一个由数十万个称为自治系统(AS)的小型网络组成的巨型网络。这些自治系统每一个本质上都是由单个组织运行的大型路由器池,自治系统是由单个组织管理的大型网络或网络组。自治系统可能有许多子网,但都共享相同的路由策略。每个自治系统都分配有自己的自治系统编号(ASN),以便轻松识别它们。

[0037] 可以理解的是,本申请实施例可以从RIPE NCC和RouteView的路由收集器获取边界网关协议的更新信息,以用于后续进行定位路由泄露的位置。

[0038] 其中,RIPE NCC为欧洲,中东及部分中亚的区域互联网注册商,为网络研究社区提供路由收集器及其收集的域间路由表和路由更新数据;RouteViews是俄勒冈大学高级网络技术中心成立的项目,为网络研究社区提供路由收集器及其收集的域间路由表和路由更新数据。

[0039] 在步骤S102中,提取更新消息中由自治系统编号组成的编码,根据编码生成自治系统三元组,并提取自治系统三元组对应的自治系统三元组特征。

[0040] 可以理解的是,本申请实施例可以从边界网关协议中的更新信息中获得自治系统编号组成的编码(AS_PATH),根据编码生成自治系统三元组,并对自治系统三元组进行特征提取。

[0041] 需要说明的是,BGP更新消息每经过一个自治系统,就会将其自治系统编号加入自己的AS_PATH中。AS_PATH中有可能出现用于路由聚合的AS_SET,该字段中包含多个自治系统编号。

[0042] 在本申请实施例中,根据编码生成自治系统三元组,包括:识别编码中预设编码、重复编号和目标字段;丢弃或移除预设编码,删除重复编码,对目标字段中包含的自治系统编号对应的自治系统生成各自对应的自治系统三元组,其中,对于满足预设条件的预设编码,丢弃编码对应的自治系统三元组。

[0043] 其中,预设编码可以为异常编码或含有环的编码等;目标字段可以为23456;预设条件是指存在预设编码、重复编码和目标字段的编码。

[0044] 可以理解的是,本申请实施例可以对获取的AS_PATH进行预处理,过滤掉非法路径及异常自治系统编号,生成自治系统三元组,具体步骤如下:

[0045] 1、丢弃含有环的AS_PATH。

[0046] 2、处理路径填充,删除AS_PATH中的重复自治系统编号,只剩下重复自治系统编号中的一个。

[0047] 3、处理AS_PATH中的AS_SET,对AS_SET中的自治系统生成各自对应的自治系统三元组。

[0048] 4、移除AS_PATH中的保留自治系统编号。其中,对于特定的自治系统编号,比如23456等,丢弃含有该号码的自治系统三元组。

[0049] 在本申请实施例中,提取自治系统三元组对应的自治系统三元组特征,包括:以自

治系统三元组为索引,查询预先建立的数据库,输出自治系统三元组特征。

[0050] 其中,数据库用于存储自治系统三元组和自治系统三元组特征之间的关系。

[0051] 可以理解的是,本申请实施例可以通过在数据库中查询自治系统三元组相应的自治系统三元组特征。

[0052] 需要说明的是,本申请需要从路由数据源周期性收集自治系统特征并存处于数据库中,路由数据源为:RIPE NCC,RouteViews,CAIDA和PeeringDB。

[0053] 其中,RIPE NCC和RouteViews在上述实施例中已经阐述,此处不再赘述。CAIDA为应用互联网数据分析中心,为网络研究社区提供互联网测量数据;PeeringDB为基于web的互联网连接信息数据库。

[0054] 在本申请实施例中,数据库存储有间隔预设时长更新的自治系统特征,其中,通过第一路由数据源更新自治系统特征中的自治系统距离、自治系统度和自治系统类型,通过第二路由数据源和第三路由数据源更新自治系统特征中的自治系统地址空间,通过第一路由数据源和第四路由数据源更新自治系统特征中的自治系统地理位置。

[0055] 其中,预设间隔时长可以依据具体情况进行设定,对此不做限定。

[0056] 可以理解的是,本申请实施例可以从路由数据源周期性收集获得自治系统特征,存储于数据库中。

[0057] 以第一路由数据源为CAIDA,第二路由数据源为RIPE NCC,第三路由数据源为RouteViews,第四路由数据源为PeeringDB为例,本申请实施例收集自治系统特征的具体步骤如下:

[0058] 1、从路由数据源CAIDA获得自治系统距离,自治系统度,自治系统类型。其中自治系统距离指的是该自治系统到各层一自治系统的平均跳数;自治系统类型分为三类,即内容提供商,企业网络和访问及传输服务提供商。本方法用0表示内容提供商,用1表示企业网络和用2表示传输服务提供商。

[0059] 2、从路由数据源RIPE NCC和RouteViews获得自治系统地址空间。

[0060] 3、从路由数据源CAIDA和PeeringDB获得自治系统地理位置。其中自治系统的地理位置指的是自治系统的所属国家和所属IXP。本方法用1-5分别表示<AS1,AS2,AS3>中3个自治系统的地理位置关系。AS1地理位置与AS2地理位置相同且二者与AS3地理位置不同,表示为1;AS1地理位置与AS3地理位置相同且二者与AS2地理位置不同,表示为2;AS2地理位置与AS3地理位置相同且二者与AS1地理位置不同,表示为3;AS1,AS2与AS3地理位置都不同表示为4;AS1,AS2与AS3地理位置都相同表示为5。

[0061] 4、将上述获得的自治系统特征存储在数据库中并周期性更新。

[0062] 需要说明的是,IXP为互联网交换中心,是互联网的物理基础设施之一,可帮助自治系统高效交换互联网流量。

[0063] 在步骤S103中,将自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器,输出路由泄露的实际位置。

[0064] 可以理解的是,本申请实施例可通过获得的自治系统三元组和对应的自治系统三元特征输入训练完成的随机森林分类器,定位路由泄露的实际位置。

[0065] 需要说明的是,路由泄露指BGP更新消息传播违背了无谷原则就会产生路由泄露,即一个自治系统不能将来自自己的提供商或对等体的路由宣告给自己的另一个提供商或

对等体。其中，无谷原则为BGP更新消息传播遵循的规则。自治系统在域间路由中是有分层高低位置之分的，一般来说，自治系统的提供商比其分层位置高，自治系统的对等体和其分层位置大体相当，自治系统的客户比其分层位置低。BGP更新消息传播所经过的自治系统所形成的路径不能形成山谷形状，即更新消息不能从高层自治系统传播到低层自治系统再传播到高层自治系统。具体来说，自治系统可以将来自自己客户的路由宣告给任何其他邻居，可以将来自自己对等体的路由宣告给任何客户邻居，可以将来自自己提供商的路由宣告给任何客户邻居。

[0066] 具体而言，将训练完成的随机森林分类器根据输入的自治系统三元组及对应的三元组特征定位路由泄露，例如定位到自治系统三元组<AS1，AS2，AS3>发生了路由泄露，那么泄露路由的自治系统为AS2，泄露路由接收者为AS1，AS2泄露的路由来自于AS3。

[0067] 在本申请实施例中，随机森林分类器的训练包括：获取路由泄露事件，其中，路由泄露事件包括发生路由泄露的自治系统三元组和未发生路由泄露的自治系统三元组；根据发生路由泄露的自治系统三元组、未发生路由泄露的自治系统三元组以及各自对应的自治系统三元组特征生成正负样本；利用正负样本对随机森林分类器进行训练，直到满足训练停止条件，得到训练完成的随机森林分类器。

[0068] 可以理解的是，本申请实施例可以从路由泄露事件数据源收集路由泄漏事件，训练随机森林分类器，从而后续使用该随机森林分类器根据自治系统三元组特征定位路由泄露事件。

[0069] 可以理解的是，随机森林分类器的训练步骤具体包括：

[0070] 1、从路由泄露事件数据源BGPStream平台收集路由泄露事件，获取发生路由泄露的自治系统三元组和正常的自治系统三元组。

[0071] 2、使用获取到的发生路由泄露的自治系统三元组和正常的自治系统三元组以及它们对应的三元组特征作为正负样本，训练随机森林分类器。

[0072] 其中，BGPStream为提供关于BGP中的劫持、泄漏和故障的告警平台。

[0073] 在本申请实施例中，在输出路由泄露的实际位置之后，还包括：生成路由泄露的报警信息，发送报警信息至预设终端。

[0074] 其中，预设终端可以为显示器等，对此不做限定。

[0075] 可以理解的是，本申请实施例可以在输出路由泄露的实际位置后，对定位到的路由泄露生成相应的报警信息，并发送到预设终端(如显示屏)，从而及时通知相应网络管理员，发现路由泄露的问题。

[0076] 下面将通过一个具体实施对路由泄露的定位方法进行阐述，如图2所示，包括以下步骤：

[0077] 步骤1、BGP更新消息预处理。从路由收集器获得BGP更新消息中的AS_PATH，并过滤掉非法路径及异常自治系统编号，生成自治系统三元组。

[0078] 步骤2、自治系统三元组特征提取。从路由数据源周期性收集自治系统特征并存储在数据库中，根据自治系统三元组生成对应的自治系统三元组特征。

[0079] 步骤3、路由泄露定位。从路由泄露事件数据源收集路由泄露事件，训练随机森林分类器，之后使用该随机森林分类器根据自治系统三元组特征定位路由泄露并生成告警，

[0080] 下面通过一个具体实施例来阐述路由泄露的定位方法，具体如下：

[0081] 1、BGP更新消息预处理。从路由收集器获得AS_PATH为[4637,4775,4766,174,3491,65000],过滤掉异常自治系统编号65000,生成自治系统三元组:<4637,4775,4766>,<4775,4766,174>和<4766,174,3491>。

[0082] 2、自治系统三元组特征提取。从路由数据源周期性收集获得自治系统特征,根据自治系统三元组生成对应的三元组特征如表1。自治系统三元组特征内容分别为:<AS1距离,AS2距离,AS3距离>,<AS1地址空间,AS2地址空间,AS3地址空间>,<AS1度,AS2度,AS3度>,<AS1类型,AS2类型,AS3类型>,3个自治系统所属国家关系和3个自治系统所属IXP关系。

[0083] 表1

[0084]

自治系统三元组	自治系统距离三元组特征	自治系统地址空间三元组特征	自治系统度三元组特征	自治系统类型三元组特征	3个自治系统所属国家关系	3个自治系统所属IXP关系
<4637,4775,4766>	<2.16,2.53,2.32>	<685824,652800,69299136>	<693,260,586>	<2,2,2>	4	5
<4775,4766,174>	<2.53,2.32,1.95>	<652800,69299136,29319936>	<260,586,6428>	<2,2,2>	4	1
<4766,174,3491>	<2.32,1.95,1.95>	<69299136,29319936,1268992>	<586,6428,743>	<2,2,2>	3	2

[0085] 3、路由泄露定位。从路由泄露事件数据源BGPStream收集路由泄露事件的AS_PATH为:[53432,13994,7029,6461,37662,37204,5511,174,25818],其中37204是泄露自治系统,其将路由泄露给了自治系统37662。发生路由泄露的自治系统三元组<37662,37204,5511>作为训练用的正样本,以及正常自治系统三元组<53432,13994,7029>,<13994,7029,6461>,<7029,6461,37662>,<6461,37662,37204>,<37204,5511,174>和<5511,174,25818>作为训练用的负样本。同理,分别获得这些自治系统三元组对应的三元组特征。之后使用样本平衡后的正负样本及其对应的三元组特征训练随机森林分类器。训练好的随机森林分类器收到自治系统三元组<3320,9002,44628>,分类器根据该自治系统三元组对应的三元组特征检测出该三元组发生了路由泄露,并定位到9002为发生泄露的自治系统,其将来自自治系统44628的路由泄露给了自治系统3320,最后生成相应的告警并及时通知相关网络管理员。当实现环境为配备英特尔至强(R)Gold 6230RCPU@2.10GHz和128GB RAM的戴尔PowerEdge R740机架式服务器时,本申请实施例可以在平均7.86毫秒内定位路由泄露,如图3所示,定位路由泄露的准确率如表2所示,可以实现90%以上的路由泄露定位准确率。

[0086] 表2

[0087]

	假阳率	召回率	精确率	准确率	F1值
本申请	0.11	0.92	0.90	0.91	0.91

[0088] 根据本申请实施例提出的路由泄露的定位方法，通过获得自治系统三元组特征以及训练随机森林分类器从BGP更新消息中的AS_PATH定位路由泄露；从单条BGP更新消息中提取AS_PATH，生成多个自治系统三元组，然后对逐个自治系统三元组判断其是否发生路由泄露；自治系统三元组特征能够准确区分路由泄露和正常路由；使用相对稳定，不需要频繁收集的自治系统三元组特征，节省定位时间并基于轻量的随机森林分类器，能够实时地定位路由泄露位置。

[0089] 其次参照附图描述根据本申请实施例提出的路由泄露的定位装置。

[0090] 图4是本申请实施例的路由泄露的定位装置的方框示意图。

[0091] 如图4所示，该路由泄露的定位装置10包括：获取模块100、提取模块200和输出模块300。

[0092] 其中，获取模块100用于获取边界网关协议的更新消息；提取模块200用于提取更新消息中由自治系统编号组成的编码，根据编码生成自治系统三元组，并提取自治系统三元组对应的自治系统三元组特征；输出模块300用于将自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器，输出路由泄露的实际位置。

[0093] 在本申请实施例中，输出模块300进一步用于：获取路由泄露事件，其中，路由泄露事件包括发生路由泄露的自治系统三元组和未发生路由泄露的自治系统三元组；根据发生路由泄露的自治系统三元组、未发生路由泄露的自治系统三元组以及各自对应的自治系统三元组特征生成正负样本；利用正负样本对随机森林分类器进行训练，直到满足训练停止条件，得到训练完成的随机森林分类器。

[0094] 在本申请实施例中，提取模块200进一步用于：以自治系统三元组为索引，查询预先建立的数据库，输出自治系统三元组特征。

[0095] 在本申请实施例中，数据库存储有间隔预设时长更新的自治系统特征，其中，通过第一路由数据源更新自治系统特征中的自治系统距离、自治系统度和自治系统类型，通过第二路由数据源和第三路由数据源更新自治系统特征中的自治系统地址空间，通过第一路由数据源和第四路由数据源更新自治系统特征中的自治系统地理位置。

[0096] 在本申请实施例中，提取模块200进一步用于：识别编码中预设编码、重复编号和目标字段；丢弃或移除预设编码，删除重复编码，对目标字段中包含的自治系统编号对应的自治系统生成各自对应的自治系统三元组，其中，对于满足预设条件的预设编码，丢弃编码对应的自治系统三元组。

[0097] 在本申请实施例中，本申请的装置10还包括：发送模块。

[0098] 其中，发送模块，用于在输出路由泄露的实际位置之后，生成路由泄露的报警信息，发送报警信息至预设终端。

[0099] 需要说明的是，前述对路由泄露的定位方法实施例的解释说明也适用于该实施例的路由泄露的定位装置，此处不再赘述。

[0100] 根据本申请实施例提出的路由泄露的定位装置，通过获得自治系统三元组特征以及训练随机森林分类器从BGP更新消息中的AS_PATH定位路由泄露；从单条BGP更新消息中提取AS_PATH，生成多个自治系统三元组，然后对逐个自治系统三元组判断其是否发生路由

泄露;自治系统三元组特征能够准确区分路由泄露和正常路由;使用相对稳定,不需要频繁收集的自治系统三元组特征,节省定位时间并基于轻量的随机森林分类器,能够实时地定位路由泄露位置。

[0101] 图5为本申请实施例提供的电子设备的结构示意图。该电子设备可以包括:

[0102] 存储器501、处理器502及存储在存储器501上并可在处理器502上运行的计算机程序。

[0103] 处理器502执行程序时实现上述实施例中提供的路由泄露的定位方法。

[0104] 进一步地,电子设备还包括:

[0105] 通信接口503,用于存储器501和处理器502之间的通信。

[0106] 存储器501,用于存放可在处理器502上运行的计算机程序。

[0107] 存储器501可能包含高速RAM(Random Access Memory,随机存取存储器)存储器,也可能还包括非易失性存储器,例如至少一个磁盘存储器。

[0108] 如果存储器501、处理器502和通信接口503独立实现,则通信接口503、存储器501和处理器502可以通过总线相互连接并完成相互间的通信。总线可以是ISA(Industry Standard Architecture,工业标准体系结构)总线、PCI(Peripheral Component,外部设备互连)总线或EISA(Extended Industry Standard Architecture,扩展工业标准体系结构)总线等。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5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。

[0109] 可选的,在具体实现上,如果存储器501、处理器502及通信接口503,集成在一块芯片上实现,则存储器501、处理器502及通信接口503可以通过内部接口完成相互间的通信。

[0110] 处理器502可能是一个CPU(Central Processing Unit,中央处理器),或者是ASIC(Application Specific Integrated Circuit,特定集成电路),或者是被配置成实施本申请实施例的一个或多个集成电路。

[0111] 本申请实施例还提供一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如上的路由泄露的定位方法。

[0112] 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申请的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不是必须针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或N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。

[0113] 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”的特征可以明示或者隐含地包括至少一个该特征。在本申请的描述中,“N个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0114] 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N个用于实现定制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,并且本申请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，这应被本申请的实施例所属技术领域的技术人员所理解。

[0115] 应当理解，本申请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，N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。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，现场可编程门阵列等。

[0116] 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。

[0117] 尽管上面已经示出和描述了本申请的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本申请的限制，本领域的普通技术人员在本申请的范围内可以对上述实施例进行变化、修改、替换和变型。

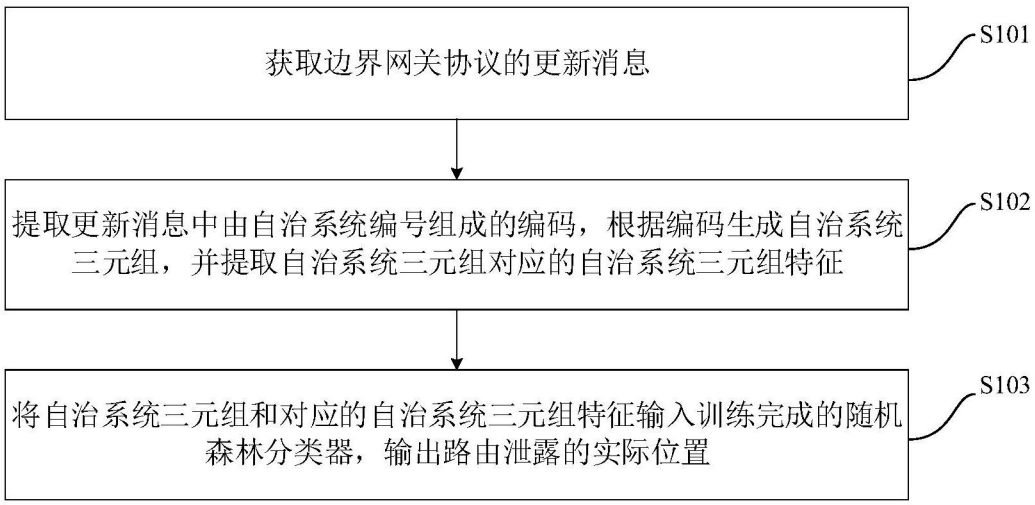

获取边界网关协议的更新消息
S101
提取更新消息中由自治系统编号组成的编码，根据编码生成自治系统三元组，并提取自治系统三元组对应的自治系统三元组特征
S102
将自治系统三元组和对应的自治系统三元组特征输入训练完成的随机森林分类器，输出路由泄露的实际位置
S103

图1

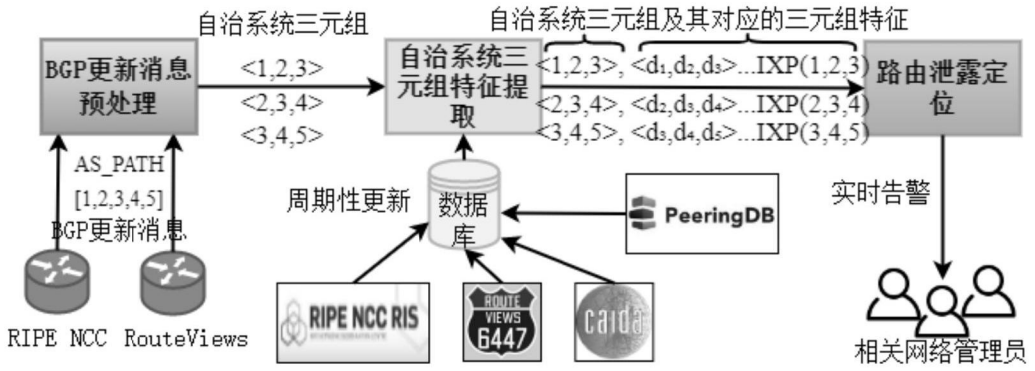

自治系统三元组
自治系统三元组及其对应的三元组特征
BGP更新消息预处理
<1,2,3>
<2,3,4>
<3,4,5>
自治系统三元组特征提取
<1,2,3>, <d1,d2,d3>...IXP(1,2,3)
<2,3,4>, <d2,d3,d4>...IXP(2,3,4)
<3,4,5>, <d3,d4,d5>...IXP(3,4,5)
路由泄露定位
AS_PATH
[1,2,3,4,5]
BGP更新消息
周期性更新
数据库
PeeringDB
实时告警
RIPE NCC
RouteViews
RIPE NCC RIS
ROUTE VIEWS 6447
caida
相关网络管理员

图2

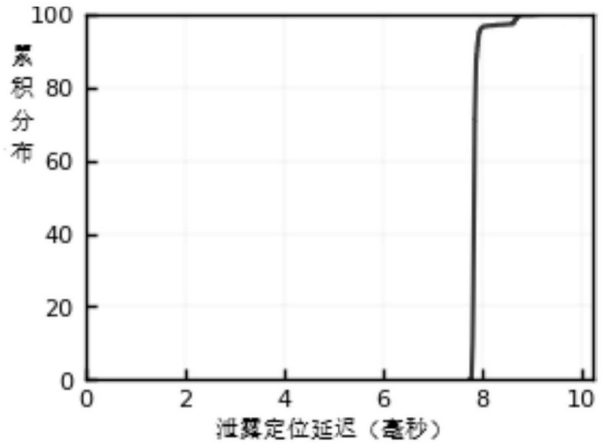

100
80
60
40
20
0
累积分布
0
2
4
6
8
10
泄露定位延迟（毫秒）

图3

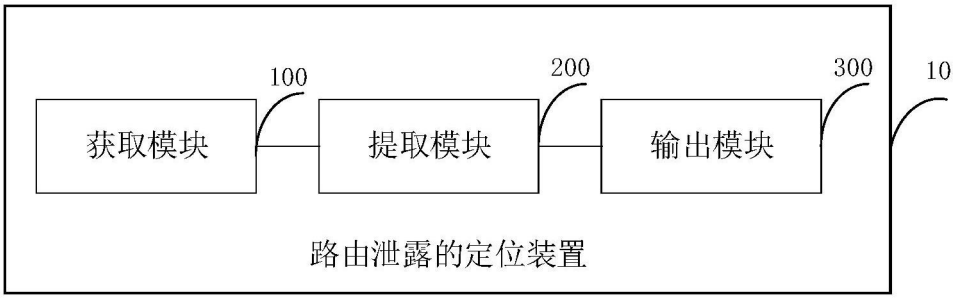

100
200
300
10
获取模块
提取模块
输出模块
路由泄露的定位装置

图4

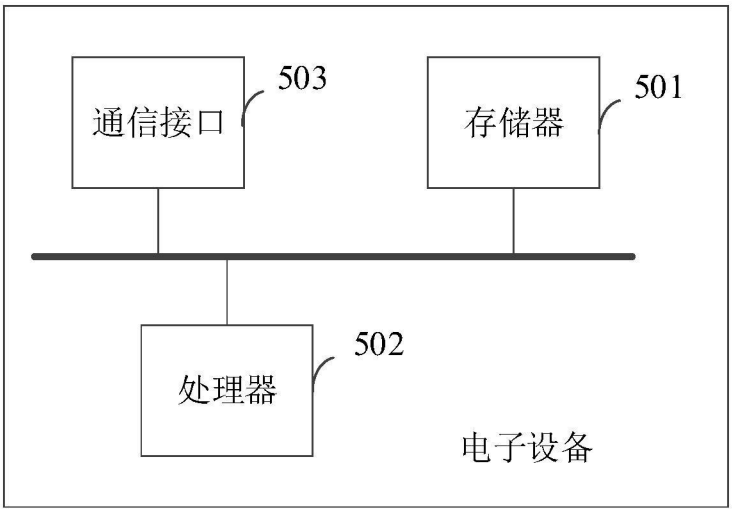

503
501
通信接口
存储器
502
处理器
电子设备

图5