(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 104184723 A
(43)申请公布日 2014.12.03

(21)申请号 201410364104.0

(22)申请日 2014.07.28

(71)申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 唐华新 严锋 舒协鎏

(74)专利代理机构 北京中博世达专利商标代理有限公司 11274
代理人 申健

(51)Int.Cl.
H04L 29/06(2006.01)
H04L 29/08(2006.01)
G06F 17/30(2006.01)

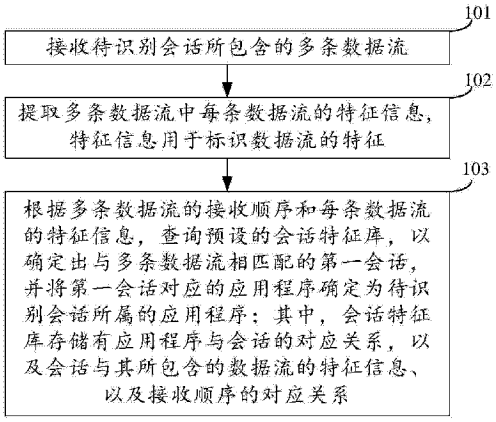
权利要求书2页 说明书14页 附图3页

(54)发明名称

一种应用程序识别方法、装置和网络设备

(57)摘要

本发明提供一种应用程序识别方法、装置和网络设备，涉及通信领域，能够解决现有技术因无法识别个别数据流，而导致的应用程序识别率低的问题，所述方法包括：接收待识别会话包含的数据流；从所述数据流中提取标识数据流特征的特征信息；在预设的会话特征库中，根据所述数据流的接收顺序和所述特征信息，确定出与所述待识别会话相匹配的第一会话，根据所述第一会话，确定出所述第一会话对应的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与数据流的特征信息、接收顺序的对应关系。本发明应用于识别数据流对应的应用程序。



CN 104184723 A

1. 一种应用程序识别方法，用于识别会话所属的应用程序，其特征在于，所述方法包括：

接收待识别会话所包含的多条数据流；

提取所述多条数据流中每条数据流的特征信息；所述特征信息用于标识数据流的特征；

根据所述多条数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，并将所述第一会话对应的应用程序确定为所述待识别会话所属的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

2. 根据权利要求 1 所述的方法，其特征在于，所述提取所述多条数据流中每条数据流的特征信息包括：

若确定所述多条数据流中的一数据流携带有标识该数据流功能的关键字符，则将所述关键字符作为该数据流的特征信息；

若确定所述多条数据流中的一数据流未携带有标识该数据流功能的关键字符，则将该数据流与所述待识别会话的前一条数据流之间的间隔流数目作为该数据流的特征信息。

3. 根据权利要求 2 所述的方法，其特征在于，所述根据所述数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，包括：

根据所述多条数据流的接收顺序，以及每条数据流的关键字符和 / 或间隔流数目，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

4. 根据权利要求 2 所述的方法，其特征在于，所述会话特征库还存储有会话所包含的各条数据流的关键字符和会话动作类型对应关系；

所述方法还包括：

根据所述待识别会话所包含的多条数据流的关键字符查询预设的会话特征库，以确定所述多条数据流中具有关键字符的数据流的会话动作类型。

5. 根据权利要求 4 所述的方法，其特征在于，所述方法还包括：

根据所述会话特征库存储的所述第一会话的各条数据流的会话动作类型，确定出所述待识别会话所包含的多条数据流中不具有关键字符的数据流的会话动作类型。

6. 根据权利要求 5 所述的方法，其特征在于，所述根据所述数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，包括：

根据所述多条数据流的接收顺序，以及每条数据流的会话动作类型，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

7. 根据权利要求 1 至 6 任意一项权利要求所述的方法，其特征在于，在接收待识别会话包含的数据流之前，所述方法还包括：

获取应用程序的会话对应的应用程序编程接口 API 在运行过程中生成的多条数据流；

确定生成所述多条数据流的接收顺序，并提取每条数据流的特征信息；

生成所述多条数据流的接收顺序和特征信息与所述会话的对应关系，并保存至所述会话特征库；生成所述应用程序与所述会话的对应关系，并保存至所述会话特征库。

8. 一种应用程序识别装置，用于识别会话所属的应用程序，其特征在于，所述应用程序识别装置包括：

基本识别单元，用于接收待识别会话所包含的多条数据流；

特征识别单元，用于提取所述多条数据流中每条数据流的特征信息；所述特征信息用于标识数据流的特征；

应用程序识别单元，用于根据所述多条数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，并将所述第一会话对应的应用程序确定为所述待识别会话所属的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

9. 根据权利要求 8 所述的装置，其特征在于，

所述特征识别单元，具体用于：若确定所述多条数据流中的一数据流携带有标识该数据流功能的关键字符，则将所述关键字符作为该数据流的特征信息；

若确定所述多条数据流中的一数据流未携带有标识该数据流功能的关键字符，则将该数据流与所述待识别会话的前一条数据流之间的间隔流数目作为该数据流的特征信息。

10. 根据权利要求 9 所述的装置，其特征在于，

在确定与所述多条数据流相匹配的第一会话的方面，所述应用程序识别单元，具体用于：根据所述多条数据流的接收顺序，以及每条数据流的关键字符和 / 或间隔流数目，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

11. 根据权利要求 9 所述的装置，其特征在于，所所述会话特征库还存储有会话所包含的各条数据流的关键字符和会话动作类型对应关系；

所述特征识别单元，还用于根据所述待识别会话所包含的多条数据流的关键字符查询预设的会话特征库，以确定所述多条数据流中具有关键字符的数据流的会话动作类型。

12. 根据权利要求 11 所述的装置，其特征在于，

所述应用程序识别单元，还用于：根据所述会话特征库存储的所述第一会话的各条数据流的会话动作类型，确定出所述待识别会话所包含的多条数据流中不具有关键字符的数据流的会话动作类型。

13. 根据权利要求 12 所述的装置，其特征在于，在确定与所述多条数据流相匹配的第一会话的方面，所述应用程序识别单元，具体用于：根据所述多条数据流的接收顺序，以及每条数据流的会话动作类型，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

14. 一种网络设备，其特征在于，所述网络设备包括权利要求 8 至 13 任一项权利要求所述的应用程序识别装置。

一种应用程序识别方法、装置和网络设备

技术领域

[0001] 本发明涉及通信领域，尤其涉及一种应用程序识别方法、装置和网络设备。

背景技术

[0002] 随着NGN(Next Generation Network，下一代网络)，3G(3rd-Generation，第三代移动通信技术)等电信技术的发展，建网成本逐步降低，导致运营商之间的竞争更加剧烈。因此，为了帮助运营商实现流量精细化运营，DPI技术(Deep Packet Inspection，中文：深度包检测)应运而出。

[0003] 现有的DPI技术是一种基于应用层的流量检测和控制技术，其识别方式一般包括：关键字符识别、端口识别、关联识别、行为识别等。DPI可以基于数据流识别出其所对应的应用程序。例如，部署有DPI的设备接收到的一条携带有应用程序类型信息的数据流后，能够根据数据流携带的应用程序类型识别出该数据流对应的应用程序类型，以此类推，部署有DPI的设备能够接收并识别出各条数据流对应的应用程序类型，进而可以拼凑出用户使用的应用程序。运营商据此能够为用户提供更为精细的服务。

[0004] 但是，现有的一些应用程序在会话过程中，会使用未携带有该应用程序类型的数据流，例如用到的基础协议数据流或加密的一些数据流等。此时，现有的DPI技术便无法根据这些数据流识别出其所对应的应用程序类型，进而也就不能拼凑出完整的应用程序，导致应用程序的识别精度不高，进而影响运营商的服务质量。

发明内容

[0005] 本发明提供一种应用程序识别方法、装置和网络设备，一定程度上解决因个别数据流未携带有应用程序类型信息，而导致的应用程序无法识别的问题。

[0006] 为达到上述目的，本发明的实施例采用如下技术方案：

[0007] 第一方面，本发明实施例提供了一种应用程序识别方法，用于识别会话所属的应用程序，所述方法包括：

[0008] 接收待识别会话所包含的多条数据流；

[0009] 提取所述多条数据流中每条数据流的特征信息；所述特征信息用于标识数据流的特征；

[0010] 根据所述多条数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，并将所述第一会话对应的应用程序确定为所述待识别会话所属的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

[0011] 在第一方面的第一种可能的实现方式中，所述提取所述多条数据流中每条数据流的特征信息包括：

[0012] 若确定所述多条数据流中的一数据流携带有标识该数据流功能的关键字符，则将所述关键字符作为该数据流的特征信息；

[0013] 若确定所述多条数据流中的一数据流未携带有标识该数据流功能的关键字符，则将该数据流与所述待识别会话的前一条数据流之间的间隔流数目作为该数据流的特征信息。

[0014] 结合第一方面第一种可能的实现方式，在第二种可能的实现方式中，所述根据所述数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，包括：

[0015] 根据所述多条数据流的接收顺序，以及每条数据流的关键字符和/或间隔流数目，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0016] 结合第一方面第一种可能的实现方式，在第三种可能的实现方式中，所述会话特征库还存储有会话所包含的各条数据流的关键字符和会话动作类型对应关系；

[0017] 所述方法还包括：

[0018] 根据所述待识别会话所包含的多条数据流的关键字符查询预设的会话特征库，以确定所述多条数据流中具有关键字符的数据流的会话动作类型。

[0019] 结合第一方面第三种可能的实现方式，在第四种可能的实现方式中，所述方法还包括：

[0020] 根据所述会话特征库存储的所述第一会话的各条数据流的会话动作类型，确定出所述待识别会话所包含的多条数据流中不具有关键字符的数据流的会话动作类型。

[0021] 结合第一方面第四种可能的实现方式，在第五种可能的实现方式中，所述根据所述数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，包括：

[0022] 根据所述多条数据流的接收顺序，以及每条数据流的会话动作类型，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0023] 结合第一方面，或者第一方面第一至第五种任意一种可能的实现方式，在第六种可能的实现方式中，在接收待识别会话包含的数据流之前，所述方法还包括：

[0024] 获取应用程序的会话对应的应用程序编程接口 API 在运行过程中生成的多条数据流；

[0025] 确定生成所述多条数据流的接收顺序，并提取每条数据流的特征信息；

[0026] 生成所述多条数据流的接收顺序和特征信息与所述会话的对应关系，并保存至所述会话特征库；生成所述应用程序与所述会话的对应关系，并保存至所述会话特征库。

[0027] 第二方面，本发明实施例提供了一种应用程序识别装置，用于识别会话所属的应用程序，所述应用程序识别装置包括：

[0028] 基本识别单元，用于接收待识别会话所包含的多条数据流；

[0029] 特征识别单元，用于提取所述多条数据流中每条数据流的特征信息；所述特征信息用于标识数据流的特征；

[0030] 应用程序识别单元，用于根据所述多条数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，并将所述第一会话对应的应用程序确定为所述待识别会话所属的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

[0031] 在第二方面的第一种可能的实现方式中，

[0032] 所述特征识别单元，具体用于：若确定所述多条数据流中的一数据流携带有标识该数据流功能的关键字符，则将所述关键字符作为该数据流的特征信息；

[0033] 若确定所述多条数据流中的一数据流未携带有标识该数据流功能的关键字符，则将该数据流与所述待识别会话的前一条数据流之间的间隔流数目作为该数据流的特征信息。

[0034] 结合第二方面第一种可能的实现方式，在第二种可能的实现方式中，

[0035] 在确定与所述多条数据流相匹配的第一会话的方面，所述应用程序识别单元，具体用于：根据所述多条数据流的接收顺序，以及每条数据流的关键字符和 / 或间隔流数目，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0036] 结合第二方面第一种可能的实现方式，在第三种可能的实现方式中，所所述会话特征库还存储有会话所包含的各条数据流的关键字符和会话动作类型对应关系；

[0037] 所述特征识别单元，还用于根据所述待识别会话所包含的多条数据流的关键字符查询预设的会话特征库，以确定所述多条数据流中具有关键字符的数据流的会话动作类型。

[0038] 结合第二方面第三种可能的实现方式，在第四种可能的实现方式中，

[0039] 所述应用程序识别单元，还用于：根据所述会话特征库存储的所述第一会话的各条数据流的会话动作类型，确定出所述待识别会话所包含的多条数据流中不具有关键字符的数据流的会话动作类型。

[0040] 结合第二方面第四种可能的实现方式，在第五种可能的实现方式中，在确定与所述多条数据流相匹配的第一会话的方面，所述应用程序识别单元，具体用于：根据所述多条数据流的接收顺序，以及每条数据流的会话动作类型，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0041] 第三方面，提供一种网络设备，网络设备包括以上任意一种应用程序识别装置。

[0042] 相较于现有技术，本发明实施例提供的应用程序识别方法、装置和网络设备不再针对单条数据流进行判断识别，而是根据待识别会话包含的所有数据流进行判断识别。这样，即使待识别会话包含的某一条数据流中没有携带应用程序类型信息，它也不会改变其在会话中所有数据流中的接收顺序，以及自身与其他数据流的特征信息，而本发明实施例正是根据会话中所有数据流的接收顺序及特征信息确定待识别会话的，因此个别没有携带应用程序类型信息的数据流不会影响本发明实施例的应用程序识别过程，提高了应用程序的识别率。

附图说明

[0043] 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。

[0044] 图 1 为本发明实施例提供的一种应用程序识别方法的流程图；

[0045] 图 2 为本发明实施例提供的另一种应用程序识别方法的流程图；

[0046] 图 3 为本发明实施例提供的再一种应用程序识别方法的流程图；

[0047] 图 4 为本发明实施例提供的一种应用程序识别装置的结构示意图；

[0048] 图 5 为本发明实施例提供的又一种应用程序识别装置的结构示意图；

[0049] 图 6 为本发明实施例提供的一种网络设备的结构示意图。

具体实施方式

[0050] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0051] 实施例一

[0052] 本发明实施例提供一种应用程序识别方法，可以应用在网关设备或其他需要对网络流量进行识别的设备上，比如 GGSN(Gateway General Packet Radio Service Support Node，网关通用分组无线服务支持节点)\P-GW(Packet data network Gateway，分组数据网网关)。应用程序运行过程中会产生多个会话，会话的每条数据流按产生的先后接收顺序会被发送至部署有本方法的网关设备上，网管设备通过抓取待识别会话中包含的数据流，识别出待识别会话所属的应用程序，即该待识别会话是由何种应用程序产生的。此外，数据流可以是五元组流形式。该五元组流是指包括五元组的数据流，五元组是由源 IP 地址，源端口，目的 IP 地址，目的端口，和传输层协议号这五个量组成的一个集合。如图 1 所示，该方法可以包括：

[0053] 步骤 101、接收待识别会话包含的多条数据流。

[0054] 步骤 102、提取多条数据流中每条数据流的特征信息，特征信息用于标识数据流的特征。

[0055] 优选地，本步骤中的特征信息可以包括：关键字符或间隔流数目，若数据流具有关键字符，则特征信息为关键字符，若数据流不具有关键字符，则特征信息为间隔流数目。

[0056] 步骤 103、根据多条数据流的接收顺序和每条数据流的特征信息，查询预设的会话特征库，以确定出与多条数据流相匹配的第一会话，并将第一会话对应的应用程序确定为待识别会话所属的应用程序；其中，会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

[0057] 相较于现有技术，本发明实施例提供的应用程序识别方法不再针对单条数据流进行判断识别，而是根据待识别会话包含的所有数据流进行判断识别。这样，即使待识别会话包含的某一条数据流中没有携带应用程序类型信息，它也不会改变其在会话中所有数据流中的接收顺序，以及自身与其他数据流的特征信息，而本发明实施例正是根据会话中所有数据流的接收顺序及特征信息确定待识别会话的，因此个别没有携带应用程序类型信息的数据流不会影响本发明实施例的应用程序识别过程，提高了应用程序的识别率。

[0058] 优选地，在本发明优选实施例中，步骤 102 可以包括：若确定多条数据流中的一数据流携带有标识该数据流功能的关键字符，则将关键字符作为该数据流的特征信息；若确定多条数据流中的一数据流未携带有标识该数据流功能的关键字符，则将该数据流与待识别会话的前一条数据流之间的间隔流数目作为该数据流的特征信息。

[0059] 在此，关键字符是数据流中的能代表或标识该数据流的报文片段，提取该关键字符可以利用现有的 DPI 技术进行。

[0060] 在会话的实际传输过程中，应用程序的一个会话的各个数据流之间有可能会固定穿插入系统的其他信令数据流，或者其它会话的数据流；同时，网关设备在接收同一会话包含的两条相邻的数据流时，中间可能会接收到其它会话的数据流，此时，同一会话的数据流之间的间隔流数目则可以作为一个数据流的特征信息。示例的，当接收到 A 会话的第一条数据流 A1 之后，穿插入系统第一条数据流 B1，之后又接收到 A 会话的第二条数据流 A2，则 A 会话的第一条数据流 A1 和第二条数据流 A2 之间的间隔流数目是 1。

[0061] 优选地，在本发明优选实施例中，在预设的会话特征库中，根据所述多条数据流的接收顺序，以及每条数据流的关键字符和 / 或间隔流数目，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0062] 进一步的，会话特征库还可以存储有会话的各条数据流的关键字符和会话动作类型对应关系。步骤 102 之后，该方法还可以包括：

[0063] 根据待识别会话所包含的多条数据流的关键字符查询预设的会话特征库，以确定多条数据流中具有关键字符的数据流的会话动作类型。

[0064] 值得说明的是，会话动作类型指示了数据流在会话中的具体操作。确定出具有该关键字符的数据流的会话动作类型后，可以为后续针对该会话动作类型数据流的操作做准备。例如进行统计、管理、计费等。

[0065] 优选地，在步骤 103 之后，该方法还可以包括：

[0066] 根据所述会话特征库存储的第一会话的各条数据流的会话动作类型，确定出待识别会话所包含的多条数据流中不具有关键字符的数据流的会话动作类型。

[0067] 优选的，步骤 103 还可以包括：根据多条数据流的接收顺序，以及每条数据流的会话动作类型，查询会话特征库，以确定出与待识别会话相匹配的第一会话。

[0068] 这样，本方案就可以确定待识别会话的每个数据流的会话动作类型，操作人员就能够通过该动作类型确定该数据流能够产生的操作。例如，假设一加密数据流的会话动作类型是输入密码，由于加密，现有技术不能够从加密数据流中获取有用的信息，而通过本实施例的方法就能够确定该数据流当前能够产生的操作是输入密码，使得操作人员能够实时监测控制数据流的传输。

[0069] 进一步的，会话特征库还可以存储有多个与应用程序一一对应的应用程序信息表，应用程序信息表包括应用程序与会话的对应关系，以及会话与数据流的特征信息、接收顺序的对应关系，该应用程序列表中还保存有对应的应用程序的协议类型。在步骤 101 之后，该方法还可以包括：获取待识别会话的第一条数据流的协议类型。相应的，步骤 103 包括：在会话特征库中，根据待识别会话的第一条数据流的协议类型，确定出第一条数据流对应的应用程序信息表，在应用程序信息表中，根据待识别会话包含的数据流的接收顺序和特征信息，确定出与待识别会话相匹配的第一会话，根据第一会话，确定出第一会话对应的应用程序。

[0070] 可以看出，根据待识别会话的第一条数据流的协议类型能够从众多的应用程序信息表中选出一个或一部分应用程序信息表，大大缩小了后续匹配的范围，能够快速识别，提高了应用程序的识别速度。

[0071] 优选地，在步骤 101 之前，该方法还可以包括：会话特征库的创建过程，其包括：获取应用程序的会话对应的API(Application Programming Interface，应用程序编程接口)在运行过程中生成的多条数据流；确定生成多条数据流的接收顺序，并提取每条数据流的特征信息；生成数据流的接收顺序和特征信息与所述会话的对应关系，并保存至所述会话特征库；生成所述应用程序与所述会话的对应关系，并保存至所述会话特征库。

[0072] 实施例二

[0073] 本发明实施例提供一种应用程序识别方法，可以应用在网关设备或其他需要流量识别的设备与装置上。本发明实施例以 GGSN(Gateway General Packet Radio Service Support Node，网关通用分组无线服务支持节点)/P-GW 为例。应用程序运行过程中会产生多个会话，会话的每条数据流按产生的先后接收顺序会被发送至部署有本方法的网关设备上。此外，数据流可以是五元组流形式。

[0074] 该五元组流是指包括五元组的数据流，五元组是由源 IP 地址，源端口，目的 IP 地址，目的端口，和传输层协议号这五个量组成的一个集合。

[0075] 如图 2 所示，本实施方法包括：

[0076] 步骤 201、生成预设的会话特征库。

[0077] 本发明实施例的会话特征库中可以存储多个与应用程序一一对应的应用程序信息表，应用程序信息表包括应用程序与会话的对应关系，以及会话与数据流的特征信息、产生接收顺序的对应关系。示例的，会话特征库中可以包括微博信息表，该微博信息表可以包括与微博对应的各个会话的对应关系。例如，登录微博的会话、回复微博的会话、发表微博会的话等等，会话中又包括许多数据流的接收顺序和特征信息的对应关系。以发表微博会话为例，发表微博的会话可以包括：进入用于写微博页面、用于访问位置服务器、用于上传微博信息和用于反馈上传的四条数据流。

[0078] 生成会话特征库的方法有很多种，例如，人工法，自动法等等。

[0079] 当采用人工法生成会话特征库时，以微博为例，会话特征库开发者可以分析微博的每一项功能在运行过程中产生的会话的数据流，将该数据流按照功能的具体步骤划分会话的数据流，从而提取会话包含的每个数据流的属性，该属性可以包括数据流的接收顺序、数据流与同一会话的上一条数据流之间的间隔流数目、数据流的关键字符和数据流的会话动作类型、数据流的协议类型等等，最后，将数据流的属性归纳成该特征信息、数据流的接收顺序或数据流的会话动作类型，每个数据流的特征信息和接收顺序的获取方法都是如此。最终，根据上述关键字符和数据流的会话动作类型、数据流的协议类型等等生成会话与数据流的接收顺序和特征信息的对应关系，并保存至会话特征库中；生成了应用程序和会话的对应关系，并保存至会话特征库中。

[0080] 当采用自动法生成会话特征库时，以微博为例，可以用 PC 通过特征提取工具访问微博的官方 API 文档网站，按照 API 文档的制定格式，提取微博各个会话对应的 API 的信息表，运行信息表中的 API，获取 API 在运行过程中与服务器产生交互的所有数据流，从而获取数据流的接收顺序和特征信息。当然，还可以获取更多的数据流的其他信息，如会话动作类型、间隔流时间差、协议类型等。之后，根据数据流的接收顺序、特征信息等，生成其与会话的对应关系，保存到会话特征库中。同时，生成应用程序与会话的对应关系，并保存至会话特征库中。

[0081] 示例性的，会话特征库中可以存储有如表 1、表 2 所示的应用程序信息表。表 1 为微博信息表，微博信息表包括了应用程序名称微博、协议类型和微博的多个会话，如表 1 所示包括了登录微博会话、发表微博会话、微博评论会话，每个会话和接收顺序号、会话动作类型、关键字符和间隔流数目相对应。表 2 为社交网站信息表，该社交网站信息表包括了应用程序名称社交网站、协议类型和社交网站的周边搜索会话，如表 1 所示，会话和接收顺序号、会话动作类型、关键字符和间隔流数目相对应。该接收顺序号为接收顺序的序号。

[0082] 在此需要说明的是，本步骤的会话特征库的创建过程可以是一个独立的过程，即利用另一独立的 PC 设备进行收集、整理、对应、创建等操作，之后再将创建得到的数据存储进如本实施例的网关设备中。

[0083] 表 1

[0084]

应用程序名称	协议类型	会话	接收顺序	特征信息		
				会话动作类型	关键字符	间隔流数目

[0085]

			号			
微博	Weibo	登录微博会话	1	weibo_login（进入登录页面）	www.weibo.com/weibo_login	1
			2	weibo-password（发送帐号密码）		1
			3	weibo_home（进入首页）	www.weibo.com/u/home	2
		发表微博会话	1	weibo_write（进入写微博页面）	www.weibo.com/weibo_write	3
			2	http_google_lbs（访问位置服务器）		2
			3	https（上传微博信息）		1
			4	feedback_to_write（反馈上传微博成功与否）	www.weibo.com/write_succeed; www.weibo.com/write_failure	1
		微博评论会话	1	weibo_comment（进入评论页面）		1
			2	https（发表评论）		1
			3	feedback_to_comment（反馈微博评论与否）	www.weibo.com/comment_succeed; www.weibo.com/comment _failure	1

[0086] 表 2

[0087]

应用程序名称	协议类型	会话	接收顺序号	特征信息		
				会话动作类型	关键字符	间隔流数目
社交网站	Social	周边	1	获取地图信息	www.XXX.com/glm/mmap(url)com.social	1

[0088]

		搜索会话	2	获取社交网站信息		2
			3	搜索		1
			4	反馈用户	gs-loc.apple.com(url)	1

[0089] 步骤 202、识别待识别会话的第一条数据流的协议类型。

[0090] 本步骤中的数据流的协议类型可以分为三种，一种是应用程序的协议类型。例如，该应用程序的协议类型可以是应用程序自身的协议类型；一种是基本协议的协议类型，例如 HTTP(Hyper Text Transfer Protocol，超文本传输协议）；一种是未知协议类型，例如 GGSN/P-GW 网关不能获取的加密的数据流的协议类型。

[0091] 步骤 203、在会话特征库中，根据待识别会话的第一条数据流的协议类型，确定出第一条数据流对应的应用程序信息表。

[0092] 根据表 1、表 2 可以看出，会话特征库中的应用程序信息表与应用程序一一对应，应用程序信息表包括了应用程序与会话的对应关系，以及会话与数据流的特征信息、接收顺序等的对应关系，此外，应用程序列表中还保存有对应的应用程序的协议类型。值得说明的是，应用程序信息表中存储的协议类型是应用程序的协议类型，不存储无应用程序对应的基本协议类型。

[0093] 示例的，假设 GGSN/P-GW 网关获取待识别会话的第一个数据流的协议类型是 Weibo，根据 Weibo 查询会话特征库中每个应用程序信息表，假设会话特征库保存有两个应用程序信息表如表 1 和表 2 所示，表 1 是微博信息表，表 2 是社交网站信息表，从而根据两张信息表确定出 Weibo 对应的应用程序信息表是微博信息表，即选出表 1。

[0094] 步骤 204、获取每个数据流的特征信息和接收顺序。

[0095] GGSN/P-GW 网关可以判断数据流是否携带有标识数据流功能的关键字符。若数据流携带有关键字符，则从数据流中提取该关键字符作为数据流的特征信息；若数据流未携带有关键字，则获取该数据流与待识别会话的前一条数据流之间的间隔流数目，在根据接收到的数据流接收顺序，确定每个数据流的接收顺序。当然，即使数据流中携带了标识数据流功能的关键字符，也仍然可以获取该数据流的间隔流数目。

[0096] 示例的，以用户需要登录微博为例，用户设备登录会话如表 3 所示，分为 3 个数据

流，用户点击界面，进入登录页面，相应的产生接收顺序号 1 的数据流；在用户输入账户密码之后，产生接收顺序号 2 的数据流，接收顺序号 2 的数据流用于发送帐号密码，不能获取到关键字符；相应的，页面跳转至微博首页，产生接收顺序号 3 的数据流，该数据流用于进入首页。其中，由于接收顺序号 1 和接收顺序号 3 的数据流是明文传输，可以解析出网站域名等 L7/L7+ 层内容作为关键字符，接收顺序号 1 的数据流的关键字符是 www.weibo.com/weibo_login，接收顺序号 3 的数据流的关键字符是 www.weibo.com/u/home；由于接收顺序号 2 是密文传输，所以该数据流无法通过传统的 DPI 方法识别关键字符，因此，只能获取数据流的接收顺序号和间隔流数目。以上接收顺序号是 GGSN/P-GW 网关接收数据流的接收顺序的接收顺序号。值得说明的是，本发明实施例中获取携带有关键字符的数据流的特征信息可以包括间隔流数目、关键字等等，以使得确定对应的会话更加准确。

[0097] 表 3

[0098]

接收顺序号	特征信息	
	关键字符	间隔流数目
1	www.weibo.com/ weibo_login	1
2		1
3	www.weibo.com/u/home	2

[0099] 步骤 205、在会话特征库中，根据会话的各条数据流的关键字符和会话动作类型对应关系，确定出具有上述关键字符的数据流的会话动作类型。

[0100] 根据表 1 获取具有关键字符的数据流的会话动作类型，很明显的可以看出，只有接收顺序号为 1 和接收顺序号为 3 的数据流具有关键字符，也相应的能够获取会话动作类型。会话动作类型用于标注数据流执行的功能，如表 4 所示。

[0101] 表 4

[0102]

接收顺序号	特征信息		
	会话动作类型	关键字符	间隔流数目
1	weibo_login（进入登录页面）	www.weibo.com/weibo_login	1
2			1
3	weibo_home（进入首页）	www.weibo.com/u/home	2

[0103] 步骤 206、在预设的会话特征库中，根据数据流的接收顺序和特征信息，确定出与待识别会话相匹配的第一会话。

[0104] 以应用程序信息表是微博信息表为例，GGSN/P-GW 网关可以在微博信息表中，将待识别会话的每个数据流的特征信息、接收顺序号分别与对应的微博的每个会话的每个数据流的特征信息、接收顺序一一进行匹配；若微博的第一会话的每个数据流的特征信息、接收顺序号与待识别会话的每个数据流的特征信息、接收顺序号相匹配，则确定出与待识别会话相匹配的第一会话。

[0105] 示例的，待识别会话的接收顺序（接收顺序号）、特征信息和会话动作类型如表 3 所示，微博信息表如表 1 所示，将表 3 的接收顺序（接收顺序号）、特征信息与表 1 中每个会话的接收顺序（接收顺序号）、特征信息进行比对，确定出与待识别会话匹配的会话是登录微博会话。值得说明的是，若确定待识别会话是登录微博会话，则可以确定表 4 空缺的会话动作类型是 weibo-password（发送帐号密码），因此，确定不具有关键字符的数据流的功能是发送帐号密码，后续过程中，其他设备或者操作人员就可以清晰的了解该数据流的功能。

[0106] 步骤 207、在预设的会话特征库中，根据第一会话，确定出第一会话对应的应用程序。

[0107] 根据会话特征库存储的应用程序和会话的对应关系，确定出第一会话对应的应用程序是微博程序，待识别会话对应的应用程序是微博。

[0108] 具体的，本实施例提供的所有步骤可以总结为两部分，如图 3 所示，第一部分是现有特征识别，第二部分是应用程序识别，数据流先经过特征识别，粗略拆选出对应的应用程序信息表和关键字符，再进一步经过应用程序识别过程确定出数据流对应的应用程序。值得说明的是，会话特征库可能存储于网关中，也可能存储在第三方设备中，图 3 表示存储在第三方设备中。

[0109] 可以看出，相较于现有技术，本发明实施例提供的应用程序识别方法不再针对单条数据流进行判断识别，而是根据待识别会话包含的所有数据流进行判断识别。这样，本发明实施例正是根据会话中所有数据流的接收顺序及特征信息确定待识别会话的，因此个别没有携带应用程序类型信息的数据流不会影响本发明实施例的应用程序识别过程，提高了应用程序的识别率，并且可以确定每个数据流具体的功能，便于后续操作。

[0110] 实施例三

[0111] 本发明实施例提供一种应用程序识别装置 30，用于执行上述方法实施例描述的方法流程。需要说明的是，本发明实施例提供的另一种应用程序识别方法的流程图；

[0112] 如图 4 所示，应用程序识别装置包括：

[0113] 基本识别单元 301，用于接收待识别会话所包含的多条数据流。

[0114] 特征识别单元 302，用于提取所述多条数据流中每条数据流的特征信息；所述特征信息用于标识数据流的特征。

[0115] 应用程序识别单元 303，用于根据所述多条数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，并将所述第一会话对应的应用程序确定为所述待识别会话所属的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

[0116] 相较于现有技术，本发明实施例提供的应用程序识别方法不再针对单条数据流进行判断识别，而是根据待识别会话包含的所有数据流进行判断识别。这样，即使待识别会话包含的某一条数据流中没有携带应用程序类型信息，它也不会改变其在会话中所有数据流中的接收顺序，以及自身与其他数据流的特征信息，而本发明实施例正是根据会话中所有数据流的接收顺序及特征信息确定待识别会话的，因此个别没有携带应用程序类型信息的数据流不会影响本发明实施例的应用程序识别过程，提高了应用程序的识别率。

[0117] 在一个优选的实施例中，特征识别单元 302 具体用于：若确定一数据流中携带有标识所述数据流功能的关键字符，则从所述数据流中提取所述关键字符作为所述数据流的特征信息；若确定一数据流中未携带有标识所述数据流功能的关键字符，则从所述数据流中提取所述数据流与所述待识别会话的前一条数据流之间的间隔流数目作为所述数据流的特征信息。

[0118] 进一步地，应用程序识别单元 303，具体用于：根据所述多条数据流的接收顺序，以及每条数据流的关键字符和 / 或间隔流数目，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0119] 值得说明的是，本发明实施例中获取携带有关键字符的数据流的特征信息可以包括间隔流数目、关键字等等，以使得确定对应的会话更加准确。

[0120] 在一个较佳的实施例中，所所述会话特征库还存储有会话所包含的各条数据流的关键字符和会话动作类型对应关系；

[0121] 所述特征识别单元，还用于根据所述待识别会话所包含的多条数据流的关键字符查询预设的会话特征库，以确定所述多条数据流中具有关键字符的数据流的会话动作类型。

[0122] 进一步地，所述应用程序识别单元，还用于：根据所述会话特征库存储的所述第一会话的各条数据流的会话动作类型，确定出所述待识别会话所包含的多条数据流中不具有关键字符的数据流的会话动作类型。

[0123] 进一步地，在确定与所述多条数据流相匹配的第一会话的方面，所述应用程序识别单元，具体用于：根据所述多条数据流的接收顺序，以及每条数据流的会话动作类型，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0124] 在另一个实施例中，会话特征库还存储有多个与应用程序一一对应的应用程序信息表，应用程序信息表包括应用程序与会话的对应关系，以及会话与数据流的特征信息、接收顺序的对应关系，应用程序列表中还保存有对应的应用程序的协议类型，

[0125] 基本识别单元 301 还用于获取到数据流组内每条数据流的协议类型。

[0126] 对应的，应用程序识别单元 303 还用于：

[0127] 在所述会话特征库中，根据所述待识别会话的第一条数据流的协议类型，确定出所述第一条数据流对应的应用程序信息表，在所述应用程序信息表中，根据所述待识别会话包含的数据流的接收顺序和特征信息，确定出与所述待识别会话相匹配的第一会话，根据所述第一会话，确定出所述第一会话对应的应用程序。

[0128] 本发明实施例提供的应用程序识别方法不再针对单条数据流进行判断识别，而是根据待识别会话包含的所有数据流进行判断识别。这样，本发明实施例正是根据会话中所有数据流的接收顺序及特征信息确定待识别会话的，因此个别没有携带应用程序类型信息

的数据流不会影响本发明实施例的应用程序识别过程，提高了应用程序的识别率。

[0129] 实施例四

[0130] 本发明实施例提供一种应用程序识别装置 40，同样，可以部署在网关设备或其他需要流量识别的设备上。如图 5 所示，该应用程序识别装置 40 可以包括处理器 401、存储器 402、接收机 404 和用于进行该应用程序识别装置 40 内部各设备之间的连接的一种或组合通信总线 403，用于实现这些设备之间的连接和相互通信。

[0131] 通信总线 403 可以是工业标准体系结构 (Industry Standard Architecture，简称为 ISA) 总线、外部设备互连 (Peripheral Component，简称为 PCI) 总线或扩展工业标准体系结构 (Extended Industry Standard Architecture，简称为 EISA) 总线等。该总线 403 可以分为地址总线、数据总线、控制总线等。

[0132] 存储器 402 可以包括只读存储器和随机存取存储器，并向处理器 401 提供指令和数据。

[0133] 接收机 404 用于接收待识别会话包含的多条数据流。

[0134] 处理器 401 用于提取多条数据流中每条数据流的特征信息；该特征信息用于标识数据流的特征；

[0135] 根据多条数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，并将所述第一会话对应的应用程序确定为所述待识别会话所属的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

[0136] 相较于现有技术，本发明实施例提供的应用程序识别方法不再针对单条数据流进行判断识别，而是根据待识别会话包含的所有数据流进行判断识别。这样，即使待识别会话包含的某一条数据流中没有携带应用程序类型信息，它也不会改变其在会话中所有数据流中的接收顺序，以及自身与其他数据流的特征信息，而本发明实施例正是根据会话中所有数据流的接收顺序及特征信息确定待识别会话的，因此个别没有携带应用程序类型信息的数据流不会影响本发明实施例的应用程序识别过程，提高了应用程序的识别率。

[0137] 进一步的，处理器 401 具体用于：若确定多条数据流中一数据流中携带有标识数据流功能的关键字符，则提取关键字符作为数据流的特征信息；若确定多条数据流中一数据流未携带有标识数据流功能的关键字符，则从数据流中提取数据流与待识别会话的前一条数据流之间的间隔流数目作为数据流的特征信息。

[0138] 进一步的，处理器 401 还可以根据所述多条数据流的接收顺序，以及每条数据流的关键字符和 / 或间隔流数目，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0139] 进一步的，会话特征库还存储有会话所包含的各条数据流的关键字符和会话动作类型对应关系，处理器 401 还可以根据所述待识别会话所包含的多条数据流的关键字符查询预设的会话特征库，以确定所述多条数据流中具有关键字符的数据流的会话动作类型，根据所述会话特征库存储的所述第一会话的各条数据流的会话动作类型，确定出所述待识别会话所包含的多条数据流中不具有关键字符的数据流的会话动作类型。

[0140] 所述处理器 401 还包括：根据所述多条数据流的接收顺序，以及每条数据流的会话动作类型，查询所述会话特征库，以确定出与所述待识别会话相匹配的第一会话。

[0141] 实施例五

[0142] 本发明实施例提供一种网络设备 50，如图 6 所示，包括以上实施例提供的任意一种应用程序识别装置 501。

[0143] 应用程序识别装置 501 可以用于接收待识别会话所包含的多条数据流；提取所述多条数据流中每条数据流的特征信息；所述特征信息用于标识数据流的特征；根据所述多条数据流的接收顺序和所述每条数据流的特征信息，查询预设的会话特征库，以确定出与所述多条数据流相匹配的第一会话，并将所述第一会话对应的应用程序确定为所述待识别会话所属的应用程序；其中，所述会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系。

[0144] 需要说明的是，本发明实施例提供的应用程序识别方法步骤的先后接收顺序可以进行适当调整，步骤也可以根据情况进行相应增减，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，可轻易想到变化的方法，都应涵盖在本发明的保护范围之内，因此不再赘述。

[0145] 本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0146] 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，可轻易想到变化或替换，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。

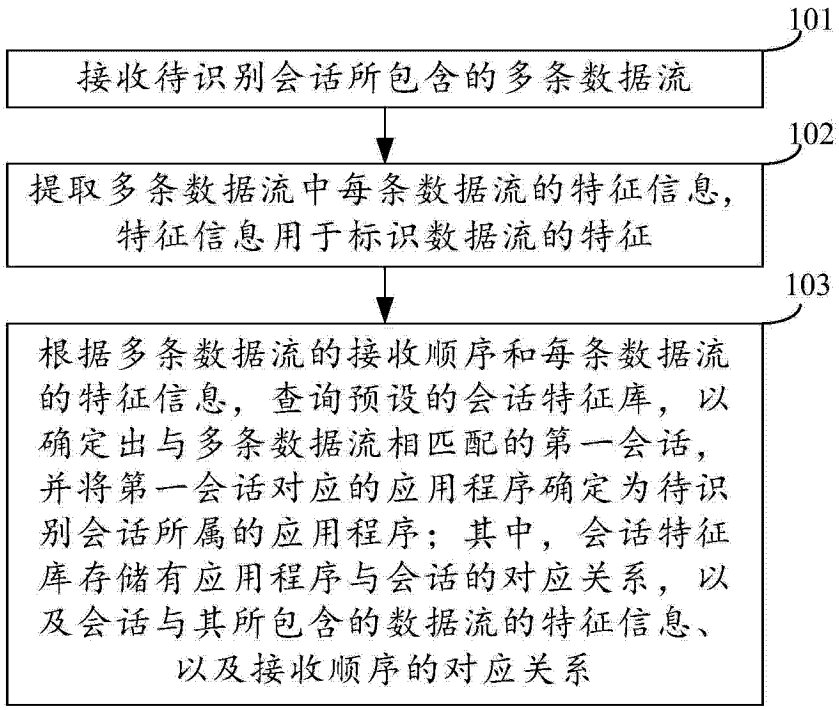

101
接收待识别会话所包含的多条数据流
102
提取多条数据流中每条数据流的特征信息，特征信息用于标识数据流的特征
103
根据多条数据流的接收顺序和每条数据流的特征信息，查询预设的会话特征库，以确定出与多条数据流相匹配的第一会话，并将第一会话对应的应用程序确定为待识别会话所属的应用程序；其中，会话特征库存储有应用程序与会话的对应关系，以及会话与其所包含的数据流的特征信息、以及接收顺序的对应关系

图 1

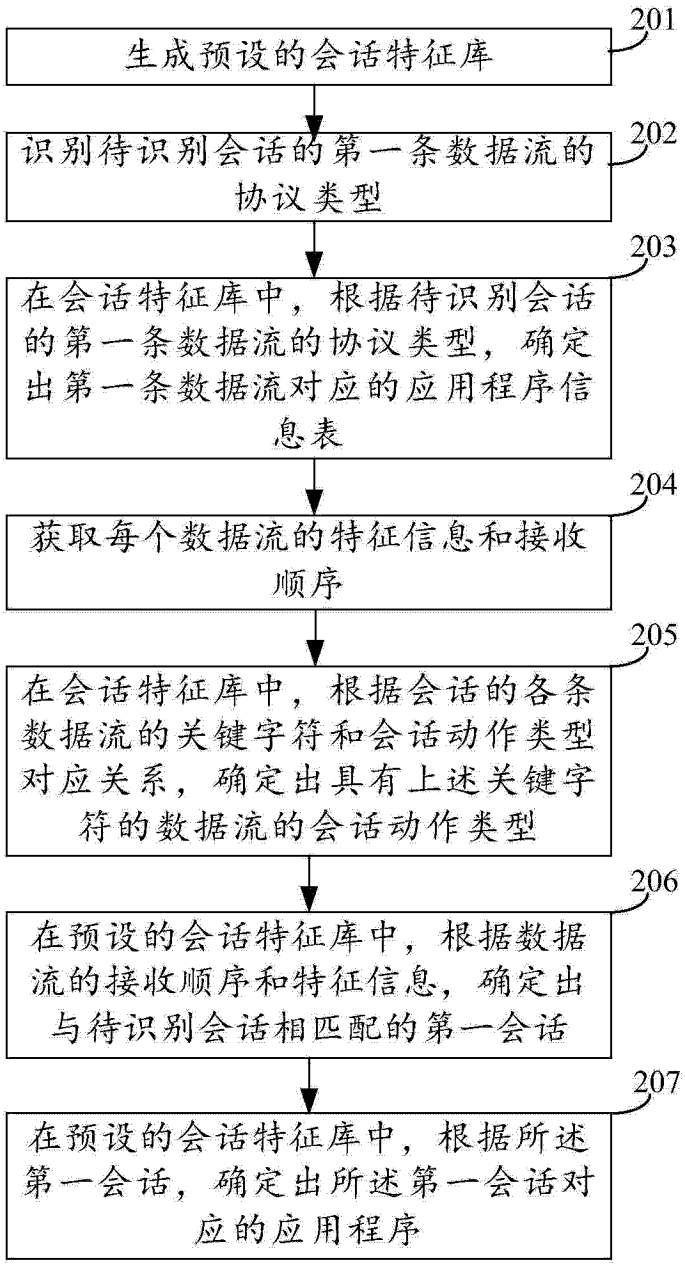

201
生成预设的会话特征库
202
识别待识别会话的第一条数据流的协议类型
203
在会话特征库中，根据待识别会话的第一条数据流的协议类型，确定出第一条数据流对应的应用程序信息表
204
获取每个数据流的特征信息和接收顺序
205
在会话特征库中，根据会话的各条数据流的关键字符和会话动作类型对应关系，确定出具有上述关键字符的数据流的会话动作类型
206
在预设的会话特征库中，根据数据流的接收顺序和特征信息，确定出与待识别会话相匹配的第一会话
207
在预设的会话特征库中，根据所述第一会话，确定出所述第一会话对应的应用程序

图 2

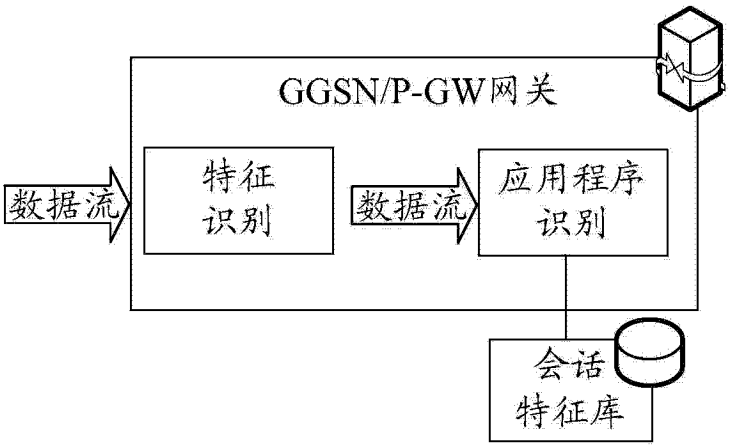

GGSN/P-GW网关
特征识别
应用程序识别
数据流
数据流
会话特征库

图 3

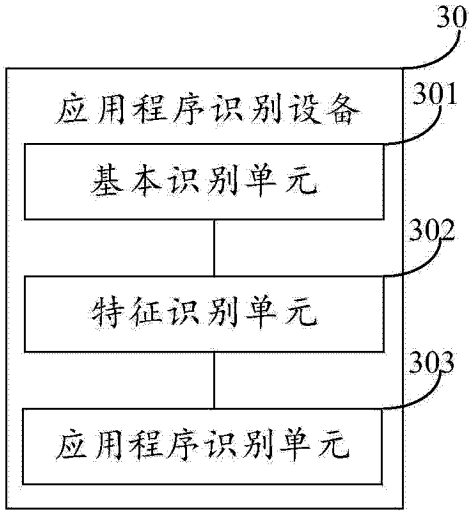

30
应用程序识别设备
301
基本识别单元
302
特征识别单元
303
应用程序识别单元

图 4

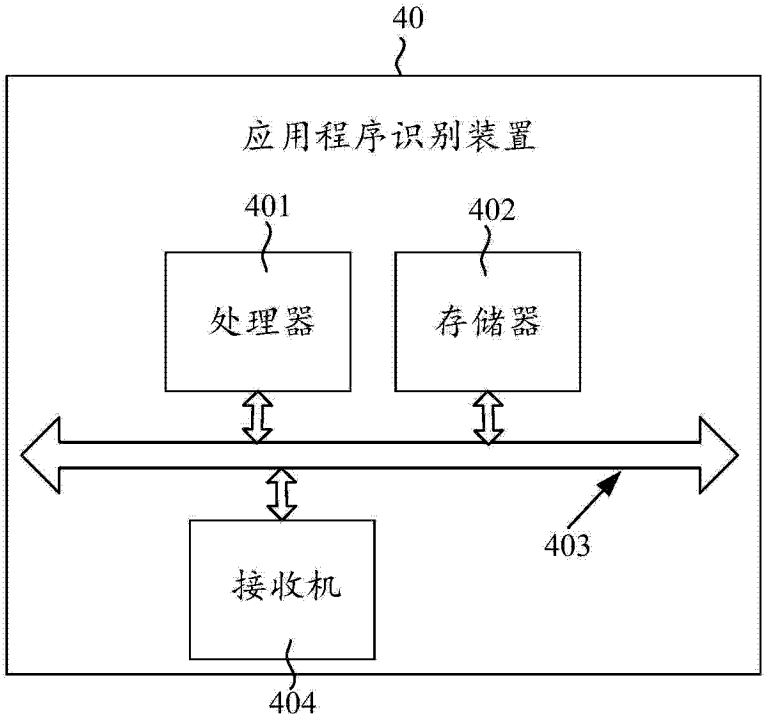

40
应用程序识别装置
401
处理器
402
存储器
403
接收机
404

图 5

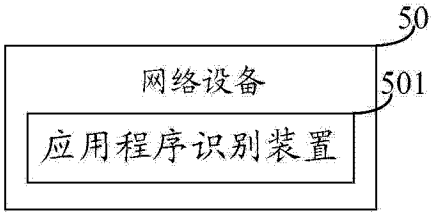

50
网络设备
501
应用程序识别装置

图 6