

一种服务质量 QoS 管理方法以及装置

5 本申请要求于 2021 年 11 月 19 日提交中国专利局、申请号为 202111373321.2、申请名称为“一种 QoS 流管理的方法”的中国专利申请以及于 2022 年 1 月 29 日提交中国专利局、申请号为 202210112669.4、申请名称为“一种服务质量 QoS 管理方法以及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

10 技术领域

本申请涉及通信领域，并且更具体地，涉及一种服务质量 QoS 管理方法以及装置。

背景技术

随着通信技术的不断发展，一些具备高速率、低时延传输需求的业务越来越普及。例如扩展现实 (extended reality, XR) 技术不断的发展和完善。XR 技术包括：虚拟现实 (virtual reality, VR)、增强现实 (augmented reality, AR) 和混合现实 (mixed reality, MR)。上述业务具有高速率、低时延传输的需求，现有的服务质量 QoS 机制无法适配上述业务的需求，例如可能造成 XR 业务产生卡顿、花屏等问题。

20 发明内容

本申请提供一种服务质量 QoS 管理方法以及装置，可以以数据包组为粒度对业务的数据流执行 QoS 控制以及对多个 QoS 流执行联合 QoS 控制，提高了 QoS 控制的灵活性，对于不同的业务有着更好的适配性。

第一方面，提供了一种 QoS 管理方法，该方法包括：第一设备发送第一信息，该第一信息用于为第一业务请求 QoS；该第一设备接收第二信息，该第二信息用于指示第一 QoS 流的 QoS 控制方式和 QoS 参数，其中该第一 QoS 流用于传输该第一业务的数据流，该 QoS 控制方式包括基于数据包组控制的第一控制方式；该第一设备根据该第二信息对该第一业务的数据流执行 QoS 控制。

本申请实施的 QoS 管理方法，第一设备可以请求基于数据包组控制的 QoS 控制方式，丰富了 QoS 管理的灵活性，可以满足不同业务的需求，提高了用户体验，降低了网络资源的浪费。

结合第一方面，在第一方面的某些实现方式中，该方法还包括：该 QoS 控制方式还包括基于数据包控制的第二控制方式。

基于上述技术方案，第一设备可以请求不同的 QoS 控制方式，从而适配不同的业务。

结合第一方面，在第一方面的某些实现方式中，该第二信息包括第一 QoS 参数，该第一 QoS 参数包括 N 个基于数据包组的 QoS 参数， $N \geq 1$ 且 N 为正整数。

基于上述技术方案，本申请实施例定义了基于数据包组的 QoS 参数，当第一设备接收到基于数据包组的 QoS 参数时，则以数据包组为粒度对第一业务的数据流执行 QoS 控

制。

结合第一方面，在第一方面的某些实现方式中，该第二信息包括第一指示信息和第二 QoS 参数，该第一指示信息用于指示该第一 QoS 流的 QoS 控制方式，该第二 QoS 参数用于指示该第一 QoS 流的 QoS 参数。

- 5 结合第一方面，在第一方面的某些实现方式中，当该第一指示信息指示该第一 QoS 流的 QoS 控制方式为第一控制方式时，该第二 QoS 参数包括至少一个基于数据包组的 QoS 参数；或当该第二指示信息指示该第一 QoS 流的 QoS 控制方式为第二控制方式时，该第二 QoS 参数包括至少一个基于数据包的 QoS 参数。

- 10 结合第一方面，在第一方面的某些实现方式中，该第二 QoS 参数包括至少一个基于数据包的 QoS 参数。

基于上述方案，本申请实施例复用了基于数据包的 QoS 参数，并通过指示信息指示 QoS 控制方式，从而第一设备根据指示信息确定 QoS 控制方式，并使用基于数据包的 QoS 参数保障 QoS 流的 QoS 需求。

- 15 结合第一方面，在第一方面的某些实现方式中，其特征在于，该第一指示信息具体用于指示该第一 QoS 流的优选 QoS 控制方式。

结合第一方面，在第一方面的某些实现方式中，该方法还包括：该第一设备接收该第一业务的业务信息，该业务信息用于确定该第一 QoS 流对应的 QoS 参数。

结合第一方面，在第一方面的某些实现方式中，该业务信息包括该第一业务的编码参数和/或该第一业务中每一个数据包组的加权因子。

- 20 结合第一方面，在第一方面的某些实现方式中，该方法还包括：该终端设备根据该第二信息和该业务信息确定该第一 QoS 流的 QoS 参数。

- 结合第一方面，在第一方面的某些实现方式中，该第二信息包括第二指示信息、第三 QoS 参数和第四 QoS 参数，该第二指示信息用于指示该第一 QoS 流的优先 QoS 控制方式，该第三 QoS 参数包括 M 个基于数据包组的 QoS 参数， $M \geq 1$ 且 M 为正整数，该第四 QoS 参数包括 L 个基于数据包的 QoS 参数， $L \geq 1$ 且 L 为正整数。

结合第一方面，在第一方面的某些实现方式中，该第一 QoS 参数包括组优先级、组延迟预算、组错误率、聚合组错误率、最大组丢失率和最大聚合组丢失率中的至少一种。

结合第一方面，在第一方面的某些实现方式中，该组优先级用于指示该第一 QoS 流中不同数据包组调度的优先级；

- 30 该组延迟预算用于指示该第一 QoS 流中数据包组在该第一设备与第二网元之间的时延上限，其中第二网元为用户面网元 UPF；该组错误率用于指示该第一 QoS 流中未成功传送的数据包组的上限；该聚合组错误率用于指示该第一 QoS 流中未成功传送的加权数据包组的上限；该最大组丢失率用于指示该第一 QoS 流中可容忍的丢弃的数据包组的上限；该最大聚合组丢失率用于指示该第一 QoS 流中可容忍的丢弃的加权数据包组的上限。

- 35 第二方面，提供了一种 QoS 管理方法，该方法包括：第一网元接收第一设备发送的第一信息，该第一信息用于为第一业务请求 QoS；该第一网元发送第二信息，该第二信息用于指示第一 QoS 流的 QoS 控制方式和 QoS 参数，其中该第一 QoS 流用于传输该第一业务的数据流，该 QoS 控制方式包括基于数据包组控制的第一控制方式。

本申请实施的 QoS 管理方法，第一网元可以根据第一设备的请求指示第一设备以数

据包组为粒度执行 QoS 控制, 丰富了 QoS 管理的灵活性, 可以满足不同业务的需求, 提高了用户体验, 降低了网络资源的浪费。

结合第二方面, 在第二方面的某些实现方式中, 该方法还包括: 该 QoS 控制方式还包括基于数据包控制的第二控制方式。

5 基于上述技术方案, 本申请实施例中包括多种不同的 QoS 控制方式, 从而可以适配不同的业务。

结合第二方面, 在第二方面的某些实现方式中, 该第二信息包括第一 QoS 参数, 该第一 QoS 参数包括 N 个基于数据包组的 QoS 参数, $N \geq 1$ 且 N 为正整数。

10 基于上述技术方案, 本申请实施例定义了基于数据包组的 QoS 参数, 当第一设备接收到基于数据包组的 QoS 参数时, 则以数据包组为粒度对第一业务的数据流执行 QoS 控制。

结合第二方面, 在第二方面的某些实现方式中, 该第二信息包括第一指示信息和第二 QoS 参数, 该第一指示信息用于指示该第一 QoS 流的 QoS 控制方式, 该第二 QoS 参数用于指示该第一 QoS 流的 QoS 参数。

15 结合第二方面, 在第二方面的某些实现方式中, 当该第一指示信息指示该第一 QoS 流的 QoS 控制方式为第一控制方式时, 该第二 QoS 参数包括至少一个基于数据包组的 QoS 参数; 或,

当该第二指示信息指示该第一 QoS 流的 QoS 控制方式为第二控制方式时, 该第二 QoS 参数包括至少一个基于数据包的 QoS 参数。

20 结合第二方面, 在第二方面的某些实现方式中, 该第二 QoS 参数包括至少一个基于数据包的 QoS 参数。

结合第二方面, 在第二方面的某些实现方式中, 该第一指示信息具体用于指示该第一 QoS 流的优选 QoS 控制方式。

25 结合第二方面, 在第二方面的某些实现方式中, 该方法还包括: 该第一网元获取该第一业务的业务信息, 该业务信息用于确定该第一 QoS 流对应的 QoS 参数。

结合第二方面, 在第二方面的某些实现方式中, 该业务信息包括该第一业务的编码参数和/或该第一业务中每一个数据包组的加权因子。

结合第二方面, 在第二方面的某些实现方式中, 该方法还包括: 该第一网元发送该业务信息。

30 结合第二方面, 在第二方面的某些实现方式中, 该第二信息包括第二指示信息、第三 QoS 参数和第四 QoS 参数, 该第二指示信息用于指示该第一 QoS 流的优选 QoS 控制方式, 该第三 QoS 参数包括 M 个基于数据包组的 QoS 参数, $M \geq 1$ 且 M 为正整数, 该第四 QoS 参数包括 L 个基于数据包的 QoS 参数, $L \geq 1$ 且 L 为正整数。

35 结合第二方面, 在第二方面的某些实现方式中, 该第一网元发送该第二信息之前, 该方法还包括:

该第一网元根据该第一业务确定该第一 QoS 流的 QoS 控制方式。

结合第二方面, 在第二方面的某些实现方式中, 该第一网元发送该第二信息之前, 该方法还包括: 该第一网元根据该第一设备是否支持第一控制方式确定该第一 QoS 流的 QoS 控制方式。

结合第二方面,在第二方面的某些实现方式中,该第一 QoS 参数包括组优先级、组延迟预算、组错误率、聚合组错误率、最大组丢失率和最大聚合组丢失率中的至少一种。

结合第二方面,在第二方面的某些实现方式中,该组优先级用于指示该第一 QoS 流中不同数据包组调度的优先级;

5 该组延迟预算用于指示该第一 QoS 流中数据包组在该第一设备与第二网元之间的时延上限,其中第二网元为用户面网元 UPF;该组错误率用于指示该第一 QoS 流中未成功传送的数据包组的上限;该聚合组错误率用于指示该第一 QoS 流中未成功传送的加权数据包组的上限;该最大组丢失率用于指示该第一 QoS 流中可容忍的丢弃的数据包组的上限;该最大聚合组丢失率用于指示该第一 QoS 流中可容忍的丢弃的加权数据包组的上限。

10 第三方面,提供了一种 QoS 管理方法,该方法包括:第一设备发送第三信息,该第三信息用于为第二业务请求 QoS;该第一设备接收第四信息,该第四信息包括用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数,其中该第二 QoS 流和第三 QoS 流用于传输该第二业务的数据流;该第一设备根据该第四信息对该第二业务的数据流执行联合 QoS 控制。

15 本申请实施的 QoS 管理方法,当第二业务的数据流承载于多个 QoS 流时,第一设备可以对该多个 QoS 流执行联合控制,提高了 QoS 管理的灵活性,满足了不同业务的需求,提高了用户体验。

结合第三方面,在第三方面的某些实现方式中,该第四信息包括第五 QoS 参数,该第五 QoS 参数包括 K 个联合 QoS 参数, $K \geq 1$ 且 K 为正整数。

20 结合第三方面,在第三方面的某些实现方式中,该第四信息包括第三指示信息和第六 QoS 参数,该第三指示信息用于指示联合控制该第二 QoS 流和该第三 QoS 流,该第六 QoS 参数为用于联合 QoS 控制的参数,该第六 QoS 包括 L 个基于数据包组的 QoS 参数, $L \geq 1$ 且 L 为正整数;或该第六 QoS 参数包括 J 个基于数据包的 QoS 参数, $J \geq 1$ 且 J 为正整数。

25 结合第三方面,在第三方面的某些实现方式中,该第五 QoS 参数包括:联合错误率、聚合联合错误率、联合最大丢失率、聚合联合最大丢失率、联合最大比特速率中的至少一种。

结合第三方面,在第三方面的某些实现方式中,该联合错误率用于指示该第二 QoS 流和该第三 QoS 流中未成功传送的数据包组或数据包的上限;

30 该聚合联合错误率用于指示该第二 QoS 流和该第三 QoS 流中未成功传送的加权数据包组或数据包的上限;该联合最大丢失率用于指示该第二 QoS 流和该第三 QoS 流中可容忍的丢弃的数据包组或数据包的上限;该聚合联合组最大丢失率用于指示该第二 QoS 流和该第三 QoS 流中可容忍的丢弃的加权数据包组或加权数据包的上限;该联合最大比特速率用于指示该第二 QoS 流和该第三 QoS 流的最大聚合比特速率的上限。

35 第四方面,提供了一种 QoS 管理方法,该方法包括:第一网元接收第一设备发送的第三信息,该第三信息用于为第二业务请求 QoS;该第一网元发送第四信息,该第四信息包括用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数,其中该第二 QoS 流和第三 QoS 流用于传输该第二业务的数据流。

本申请实施的 QoS 管理方法,当业务的数据流承载于多个 QoS 流时,第一网元可以指示第一设备对该多个 QoS 流执行联合控制,提高了 QoS 管理的灵活性,满足了不同业

务的需求,提高了用户体验。

结合第四方面,在第四方面的某些实现方式中,该第四信息包括第五 QoS 参数,该第五 QoS 参数包括 K 个联合 QoS 参数, $K \geq 1$ 且 K 为正整数。

结合第四方面,在第四方面的某些实现方式中,该第四信息包括第三指示信息和第六 QoS 参数,该第三指示信息用于指示联合控制该第二 QoS 流和该第三 QoS 流,该第六 QoS 参数为用于联合 QoS 控制的参数,该第六 QoS 包括 L 个基于数据包组的 QoS 参数, $L \geq 1$ 且 L 为正整数;或该第六 QoS 参数包括 J 个基于数据包的 QoS 参数, $J \geq 1$ 且 J 为正整数。

结合第四方面,在第四方面的某些实现方式中,该第五 QoS 参数包括:联合错误率、聚合联合错误率、联合最大丢失率、聚合联合最大丢失率、联合最大比特速率中的至少一种。

结合第四方面,在第四方面的某些实现方式中,该联合错误率用于指示该第二 QoS 流和该第三 QoS 流中未成功传送的数据包组或数据包的上限;该聚合联合错误率用于指示该第二 QoS 流和该第三 QoS 流中未成功传送的加权数据包组或数据包的上限;该联合最大丢失率用于指示该第二 QoS 流和该第三 QoS 流中可容忍的丢弃的数据包组或数据包的上限;该聚合联合最大丢失率用于指示该第二 QoS 流和该第三 QoS 流中可容忍的丢弃的加权数据包组或加权数据包的上限;该联合最大比特速率用于指示该第二 QoS 流和该第三 QoS 流的最大聚合比特速率的上限。

第五方面,提供一种通信的装置,该装置用于执行上述第一方面至第四方面任一种可能实现方式中的方法。具体地,该装置可以包括用于执行第一方面至第四方面任一种可能实现方式中的方法的单元和/或模块,如处理单元和/或通信单元。在一种实现方式中,该装置为通信设备(如终端设备,又如网络设备)。当该装置为通信设备时,通信单元可以是收发器,或,输入/输出接口;处理单元可以是至少一个处理器。可选地,收发器可以为收发电路。可选地,输入/输出接口可以为输入/输出电路。在另一种实现方式中,该装置为用于通信设备(如终端设备,又如网络设备)的芯片、芯片系统或电路。当该装置为用于通信设备的芯片、芯片系统或电路时,通信单元可以是该芯片、芯片系统或电路上的输入/输出接口、接口电路、输出电路、输入电路、管脚或相关电路等;处理单元可以是至少一个处理器、处理电路或逻辑电路等。

第六方面,提供一种通信的装置,该装置包括:至少一个处理器,用于执行存储器存储的计算机程序或指令,以执行上述第一方面至第四方面任一种可能实现方式中的方法。

可选地,该装置还包括存储器,用于存储的计算机程序或指令。可选地,该装置还包括通信接口,处理器通过通信接口读取存储器存储的计算机程序或指令。

在一种实现方式中,该装置为通信设备(如终端设备,又如网络设备)。

在另一种实现方式中,该装置为用于通信设备(如终端设备,又如网络设备)的芯片、芯片系统或电路。

第七方面,本申请提供一种处理器,用于执行上述第一方面至第四方面提供的方法。对于处理器所涉及的发送和获取/接收等操作,如果没有特殊说明,或者,如果未与其在相关描述中的实际作用或者内在逻辑相抵触,则可以理解为处理器输出和接收、输入等操作,也可以理解为由射频电路和天线所进行的发送和接收操作,本申请对此不做限定。

第八方面,提供一种计算机可读存储介质,该计算机可读介质存储用于设备执行的程

序代码,该程序代码包括用于执行上述第一方面至第四方面任一种可能实现方式中的方法。

第九方面,提供一种包含指令的计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述第一方面至第四方面任一种可能实现方式中的方法。

第十方面,提供一种通信系统,包括前述的终端设备和网络设备。

- 5 其中,第五方面至第十方面的有益效果,请参见第一方面至第四方面的有益效果,不重复赘述。

附图说明

图 1 是 5G 移动通信系统架构示意图。

- 10 图 2 是数据包与数据包组的示意图。

图 3 PDU 会话中 QoS 流的示意图。

图 4 是本申请提供的一种 QoS 管理方法的示意性流程图。

图 5 是本申请提供的一种 QoS 管理方法的示意性流程图。

图 6 是本申请提供的一种 QoS 管理方法的示意性流程图。

- 15 图 7 是本申请提供的一种 QoS 管理方法的示意性流程图。

图 8 是本申请提供的一种 QoS 管理方法的示意性流程图。

图 9 本申请提供的一种装置的示意性框图。

图 10 本申请提供的一种装置的示意性框图。

20 具体实施方式

下面将结合附图,对本申请中的技术方案进行描述。

- 本申请实施例的技术方案可以应用于各种通信系统,例如:全球移动通讯(Global System of Mobile communication, GSM)系统、码分多址(Code Division Multiple Access, CDMA)系统、宽带码分多址(Wideband Code Division Multiple Access, WCDMA)系统、
25 通用分组无线业务(General Packet Radio Service, GPRS)、长期演进(Long Term Evolution, LTE)系统、LTE 频分双工(Frequency Division Duplex, FDD)系统、LTE 时分双工(Time Division Duplex, TDD)、通用移动通信系统(Universal Mobile Telecommunication System, UMTS)、全球互联微波接入(Worldwide Interoperability for Microwave Access, WiMAX)通信系统、第五代(5th Generation, 5G)系统或新无线(New Radio, NR),以及未来演进的通信系统等。
30

- 图 1 示出了本申请实施例适用的一种通信系统的网络架构示意图,该网络架构中包括终端设备、接入网设备、接入管理网元、会话管理网元、用户面网元、策略控制网元、网络切片选择网元、网络仓库功能网元、网络数据分析网元、统一数据管理网元、统一数据存储网元、认证服务功能网元、网络能力开放网元、应用功能网元,以及连接运营商网络
35 的数据网络(data network, DN)。终端设备可通过接入网设备、用户面网元向数据网络发送业务数据,以及从数据网络接收业务数据。

其中,终端设备是一种具有无线收发功能的设备,可以部署在陆地上,包括室内或室外、手持、穿戴或车载;也可以部署在水面上(如轮船等);还可以部署在空中(例如飞机、气球和卫星上等)。所述终端设备可以经无线接入网(radio access network, RAN)与

核心网进行通信,与 RAN 交换语音和/或数据。所述终端设备可以是手机(mobile phone)、平板电脑(Pad)、带无线收发功能的电脑、移动互联网设备(mobile internet device, MID)、可穿戴设备、虚拟现实(virtual reality, VR)终端设备、增强现实(augmented reality, AR)终端设备、工业控制(industrial control)中的无线终端、无人驾驶(self driving)中的无线终端、远程医疗(remote medical)中的无线终端、智能电网(smart grid)中的无线终端、运输安全(transportation safety)中的无线终端、智慧城市(smart city)中的无线终端、智慧家庭(smart home)中的无线终端、无人机、无人机控制器等等。本申请的实施例对应用场景不做限定。终端设备有时也可以称为用户设备(user equipment, UE)、移动台和远方站等,本申请的实施例对终端设备所采用的具体技术、设备形态以及名称不做限定。

10 接入网设备,是网络中用于将终端设备接入到无线网络的设备。所述接入网设备可以为无线接入网中的节点,又可以称为基站,还可以称为无线接入网(radio access network, RAN)节点(或设备)。网络设备可以包括长期演进(long term evolution, LTE)系统或演进的 LTE 系统(LTE-Advanced, LTE-A)中的演进型基站(NodeB 或 eNB 或 e-NodeB, evolutionary Node B),如传统的宏基站 eNB 和异构网络场景下的微基站 eNB,或者也可以包括第五代移动通信技术(5th generation, 5G)新无线(new radio, NR)系统中的下一代节点 B(next generation node B, gNB),或者还可以包括无线网络控制器(radio network controller, RNC)、节点 B(Node B, NB)、基站控制器(base station controller, BSC)、基站收发台(base transceiver station, BTS)、传输接收点(transmission reception point, TRP)、家庭基站(例如,home evolved NodeB,或 home Node B, HNB)、基带单元(base band unit, BBU)、基带池 BBU pool,或 WiFi 接入点(access point, AP)等,再或者还可以包括云接入网(cloud radio access network, CloudRAN)系统中的集中式单元(centralized unit, CU)和分布式单元(distributed unit, DU),本申请实施例并不限定。在接入网设备包括 CU 和 DU 的分离部署场景中, CU 支持无线资源控制(radio resource control, RRC)、分组数据汇聚协议(packet data convergence protocol, PDCP)、业务数据适配协议(service data adaptation protocol, SDAP)等协议; DU 主要支持无线链路控制层(radio link control, RLC)、媒体接入控制层(media access control, MAC)和物理层协议。

接入管理网元,主要用于移动网络中的终端的附着、移动性管理、跟踪区更新流程,接入管理网元终结了非接入层(non access stratum, NAS)消息、完成注册管理、连接管理以及可达性管理、分配跟踪区域列表(track area list, TA list)以及移动性管理等,并且透明路由会话管理(session management, SM)消息到会话管理网元。在第五代(5th generation, 5G)通信系统中,接入管理网元可以是接入与移动性管理功能(access and mobility management function, AMF),在未来的通信系统(如 6G 通信系统)中,移动性管理网元可以仍是 AMF 网元,或者也可以具有其它名称,本申请并不限定。

35 会话管理网元,主要用于移动网络中的会话管理,如会话建立、修改、释放。具体功能如为终端分配互联网协议(internet protocol, IP)地址、选择提供报文转发功能的用户面网元等。在 5G 通信系统中,会话管理网元可以是会话管理功能(session management function, SMF),在未来的通信系统(如 6G 通信系统)中,会话管理网元可以仍是 SMF 网元,或者也可以具有其它名称,本申请并不限定。

用户面网元，主要用于对用户报文进行处理，如转发、计费、合法监听等。用户面网元也可以称为协议数据单元(protocol data unit, PDU)会话锚点(PDU session anchor, PSA)。在 5G 通信系统中，用户面网元可以是用户面功能(user plane function, UPF)，在未来的通信系统(如 6G 通信系统)中，用户面网元可以仍是 UPF 网元，或者也可以具有其它名称，本申请并不限定。

策略控制网元，包含用户签约数据管理功能、策略控制功能、计费策略控制功能、服务质量(quality of service, QoS)控制等。在 5G 通信系统中，策略控制网元可以是策略控制功能(policy control function, PCF)，在未来的通信系统(如 6G 通信系统)中，策略控制网元可以仍是 PCF 网元，或者也可以具有其它名称，本申请并不限定。

网络切片选择功能网元，主要用于为终端设备的业务选择合适的网络切片。在 5G 通信系统中，网络切片选择网元可以是网络切片选择功能(network slice selection function, NSSF)网元，在未来的通信系统(如 6G 通信系统)中，网络切片选择网元可以仍是 NSSF 网元，或者也可以具有其它名称，本申请并不限定。

网络仓库功能网元，主要用于提供网元或网元所提供服务的注册和发现功能。在 5G 通信系统中，网络仓库功能网元可以是网络仓库功能(network repository function, NRF)，在未来的通信系统(如 6G 通信系统)中，网络仓库功能网元可以仍是 NRF 网元，或者也可以具有其它名称，本申请并不限定。

网络数据分析网元，可以从各个网络功能(network function, NF)，例如策略控制网元、会话管理网元、用户面网元、接入管理网元、应用功能网元(通过网络能力开放功能网元)收集数据，并进行分析和预测。在 5G 通信系统中，网络数据分析网元可以是网络数据分析功能(network data analytics function, NWDAF)，在未来的通信系统(如 6G 通信系统)中，网络数据分析网元可以仍是 NWDAF 网元，或者也可以具有其它名称，本申请并不限定。

统一数据管理网元，主要用于管理终端设备的签约信息。在 5G 通信系统中，统一数据管理网元可以是统一数据管理(unified data management, UDM)，在未来的通信系统(如 6G 通信系统)中，统一数据管理网元可以仍是 UDM 网元，或者也可以具有其它名称，本申请并不限定。

统一数据存储网元，主要用于存储结构化的数据信息，其中包括签约信息、策略信息，以及有标准格式定义的网络数据或业务数据。在 5G 通信系统中，统一数据存储网元可以是统一数据存储(unified data repository, UDR)，在未来的通信系统(如 6G 通信系统)中，统一数据存储网元可以仍是 UDR 网元，或者也可以具有其它名称，本申请并不限定。

认证服务功能网元，主要用于对终端设备进行安全认证。在 5G 通信系统中，认证服务功能网元可以是认证服务器功能(authentication server function, AUSF)，在未来的通信系统(如 6G 通信系统)中，认证服务功能网元可以仍是 AUSF 网元，或者也可以具有其它名称，本申请并不限定。

网络能力开放网元，可以将网络的部分功能有控制地暴露给应用。在 5G 通信系统中，网络能力开放网元可以是网络能力开放功能(network exposure function, NEF)，在未来的通信系统(如 6G 通信系统)中，网络能力开放网元可以仍是 NEF 网元，或者也可以具有其它名称，本申请并不限定。

应用功能网元，可以向运营商的通信网络的控制面网元提供各类应用的服务数据，或者从通信网络的控制面网元获得网络的数据信息和控制信息。在 5G 通信系统中，应用功能网元可以是应用功能(application function, AF)，在未来的通信系统(如 6G 通信系统)中，应用功能网元可以仍是 AF 网元，或者也可以具有其它名称，本申请并不限定。

5 数据网络，主要用于为终端设备提供数据传输服务。数据网络可以是私有网络，如局域网，也可以是公用数据网(public data network, PDN)网络，如因特网(Internet)，还可以是运营商共同部署的专有网络，如配置的 IP 多媒体网络子系统(IP multimedia core network subsystem, IMS)服务。

10 应理解，上述网元或者功能既可以是硬件设备中的网络元件，也可以是在专用硬件上运行的软件功能，或者是平台(例如，云平台)上实例化的虚拟化功能。可选的，上述网元或者功能可以由一个设备实现，也可以由多个设备共同实现，还可以是一个设备内的一个功能模块，本申请实施例对此不作具体限定。

以下，为便于理解本申请提出的技术方案，对本申请实施例中涉及的应用中的相关技术概念进行说明。

15 1、XR 技术是指通过计算机技术和可穿戴设备产生的一个真实与虚拟组合、可人机交互的环境。XR 技术具有多视角、交互性强等优点，能够为用户提供了一种全新的体验，具有极大的应用价值和商业潜力。XR 包含 VR、AR 和 MR 等技术，能够广泛应用于娱乐、游戏、医疗、广告、工业、在线教育、以及工程等诸多领域。VR 技术主要是指对视觉和音频场景的渲染以尽可能地模拟现实世界中的视觉和音频对用户的感官刺激，VR 技术通常要求用户佩戴 XR 终端(例如头戴式设备)进而向用户模拟视觉和/或听觉。VR 技术还可以对用户进行动作跟踪，从而及时更新模拟的视觉和/或听觉内容。AR 技术主要是指在用户感知的现实环境中提供视觉和/或听觉的附加信息或人工生成内容，其中，用户对现实环境的获取可以是直接的(例如不进行感测、处理和渲染)，也可以是间接的(例如通过传感器等方式进行传递)，并进行进一步的增强处理。MR 技术是将一些虚拟元素插入到物理场景

20 中，目的是为用户提供一种这些元素是真实场景一部分的沉浸体验。网络设备可以对 XR 业务产生的数据(可称为 XR 数据)进行处理和传输，例如云端的网络设备可以对 XR 的源数据进行渲染和编码(比如信源编码)，借助核心网和/或接入网的网络设备将 XR 数据传输到 XR 终端。XR 终端通过对 XR 数据的处理为用户提供多样化的 XR 体验(例如沉浸体验、视觉体验、交互体验或设备体验等)。XR 体验有多种不同的评价维度，例如包括以下

25 评价维度中的一种或多种：画面清晰度、画面流畅度、画面畸变、画面立体感、画面黑边、画面拖影、音质、音效、视场角、卡顿感、花屏感、眩晕感、音视频同步、交互自由度、交互操作响应速度、交互操作精准度、交互内容加载速度等。

30 对于 XR 等媒体类的业务，在执行媒体编码时，(例如 H.263, H.264, MPEG4 等)，其业务数据流通常由一系列的媒体单元(media unit, MU)组成，每个媒体单元代表了应用层的一个完整的数据单元，包括一个或多个数据包(packet)如 IP 包，这些数据包之间具有强相关性，具有完整性的传输需求，当其中一个或多个数据包传输失败时，接收端可能无法解码整个媒体单元。在一些实施例中，媒体单元也可以被称为数据包组、帧(例如关键帧 I 帧、前向差别帧 P 帧、双向差别帧 B 帧等)、数据分片、应用层数据单元、数据单元等。例如，如图 2 所示，第一数据包组包括第一数据包、第二数据包、第三数据包、

第四数据包, 构成一个媒体单元。当上述数据包中的一个或多个数据包传输失败时, 接收端可能无法解码第一数据包组, 从而影响 XR 业务的体验。

进一步地, XR 业务的多个数据包组之间可能存在关联, 当一个数据包组解码失败时, 可能会造成其他数据包组也解码失败, 例如 I 帧解码失败, 可能会导致后续的几个 P 帧、B 帧均解码失败。

2、服务质量 (Quality of Service, QoS) 是指用来解决网络延迟和阻塞等问题的一种技术。当网络发生拥塞的时候, 数据有可能被丢弃, 为满足用户对不同应用不同 QoS 的要求, 就需要网络能根据用户的要求分配和调度资源, 对不同的数据提供不同的 QoS。在 5G 系统中, 采用了基于流的 QoS 模型。映射到相同 QoS 流的数据会受到相同的转发处理。

10 如图 3 所示, UE 和 UPF 之间建立有 PDU 会话, UE 和 NB 之间的为无线承载, NB 和 UPF 之间为核心网隧道, PDU 会话中包括多个 QoS 流, 该多个 QoS 流包括第一 QoS 流、第二 QoS 流等。该多个 QoS 流可以用于传输不同业务的数据流, 也可以用于传输同一个业务的数据流。例如, 第一 QoS 流用于传输第一业务的数据流, 第二 QoS 流用于传输第二业务的数据流, 或第一 QoS 流用于传输第一业务的第一数据流, 第二 QoS 流用于传输第一业务

15 业务的第二数据流。

3、QoS 参数是指 QoS 关联的 QoS 需求, 包括但不限于 QoS 参数 (QoS parameters)、QoS 特性 (QoS Characteristics) 等。QoS 参数也可以称为 QoS 流级别 QoS 参数 (QoS Flow level QoS parameters)、QoS 参考 (QoS reference)、QoS 流描述 (QoS flow descriptions)、QoS 信息 (QoS Information) 等。换句话说, 通过 QoS 参数可以量化体现 QoS 需求。本

20 申请实施例中, QoS 参数不仅包括基于数据包组的 QoS 参数还新定义了基于数据包组的 QoS 参数、联合 QoS 参数、混合 QoS 参数。

基于数据包的 QoS 参数可能包括下述中的一个或多个: 5G QoS 标识 (5G QoS Identifier, 5QI)、分配保持优先级 (allocation and retention priority, ARP)、反射 QoS 属性 (Reflective QoS Attribute, RQA)、通知控制 (Notification control)、流比特率 (Flow Bit Rates, FBR)、聚合比特率 (Aggregate Bit Rates, ABR)、默认值 (Default values)、最大包丢失率 (Maximum Packet Loss Rate, MPLR)、资源类型 (Resource Type)、优先级 (Priority Level, PL)、包延迟预算 (Packet Delay Budget, PDB)、包错误率 (Packet Error Rate, PER)、平均窗口 (Averaging Window)、最大数据突发量 (Maximum Data Burst Volume, MDBV) 等, 其中资源类型包括保证流比特 (Guaranteed Bit Rate, GBR) 和非保证流比特 (NonGuaranteed Bit Rate, Non-GBR)。基于数据包的 QoS 参数为目前使用的 QoS 参数, 在此不再详细介绍。

25

30

基于数据包组的 QoS 参数、联合 QoS 参数、混合 QoS 参数的详细介绍请参见后文、

目前, 由于是以数据包为粒度控制 QoS 流承载的数据, 无法满足一些需要高速率、低时延传输需求的业务。以 XR 业务为例, 例如 XR 业务产生的数据包括 3 个数据包组, 该 3 个数据包组分别为第一数据包组、第二数据包组、第三数据包组, 每一个数据包组包

35 括 100 个数据包, 且第一数据包组和第二数据包组之间存在关联, 第一数据包组和第三数据包组之间存在关联, 即当第一数据包组无法解码时, 第二数据包组、第三数据包组也无法解码。第一 QoS 流用于传输该 XR 业务, 且该第一 QoS 流对应的 QoS 参数包括数据包错误率 (packet error rate, PER)。假设 PER 为 0.03, 则可以理解为第一 QoS 流在传输该

XR 业务时传输错误的数据包的上限为 9 个, 换句话说, 仅需要保证第一 QoS 流在传输该 XR 业务时传输错误的数据包低于 9 个就可以满足第一 QoS 流的 QoS 需求。但是, 由于第一数据包组和第二数据包组之间存在关联, 第一数据包组和第三数据包组之间存在关联, 假设第一数据包组中的数据包发送传输错误, 则接收端无法解码第一数据包组, 即使在第一 QoS 流传输错误的数据包的个数低于 9 个, 在接收端还是无法解码上述三个数据包组, 从而造成了网络资源的浪费, 影响用户体验。

综上所述, 目前基于数据包为粒度控制 QoS 流承载的数据的方法无法满足一些具备高速率、低时延传输需求的业务。基于此, 本申请提出了一种 QoS 管理方法, 可以以数据包组为粒度控制 QoS 流, 可以避免网络资源的浪费, 提升用户体验。

图 4 示出了本申请提供的一种 QoS 管理方法 400 的示意性流程图。

S401, 第一设备向第一网元发送第一信息, 该第一信息用于为第一业务请求 QoS。

相应的, 第一网元接收第一设备发送的第一信息。

具体地, 第一设备向第一网元发送第一信息, 该第一信息用于为第一业务请求 QoS。

一种可能的实现方式, 该第一信息包括第一业务的标识 (例如, 应用程序标识等)、五元组 (源 IP 地址、源端口号、目标 IP 地址、目标端口号、传输层协议)、三元组 (目标 IP 地址、目标端口号、传输层协议) 中的至少一个以及请求信息, 该请求信息用于为第一业务请求 QoS。

可选的, 在另一种可能的实现方式中, 该请求信息还用于请求 QoS 流的控制方式。QoS 流的控制方式包括基于数据包组控制的第一控制方式和基于数据包控制的第二控制方式。基于数据包组控制的第一控制方式可以理解为以数据包组为粒度执行 QoS 控制。例如, QoS 流传输有 100 个数据包, 该 100 个数据包前 50 个数据包为第一数据包组, 后 50 个数据包为第二数据包组, 当需要丢弃数据时, 若执行的是第一控制方式, 则会以数据包组为粒度丢弃第一数据包组和/或第二数据包组, 若执行的是第二控制方式, 则会以数据包为粒度丢弃该 100 个数据包中的一个或多个数据包。为了便于说明, 下文中以第一控制方式为基于数据包组控制, 第二控制方式为基于数据包控制进行介绍。

该请求信息可以直接请求 QoS 流的控制方式或间接请求 QoS 流的控制方式。

在一些实施例中, 该请求信息可以间接请求 QoS 流的控制方式。例如, 该请求信息请求的是 QoS 参数。可以理解的是, 当该请求信息请求的是基于数据包组的 QoS 参数, 则该请求信息请求的 QoS 控制方式为第一控制方式。类似地, 当该请求信息请求的是基于数据包的 QoS 参数, 则该请求信息请求的 QoS 控制方式为第二控制方式。当该请求信息请求的是混合 QoS 参数, 则可以根据该请求信息请求的混合 QoS 参数确定 QoS 控制方式 (例如通过混合 QoS 参数的 5QI 取值等, 具体请参见后文说明)。

应理解, 基于数据包组的 QoS 参数和混合 QoS 参数为本申请实施例提出的 2 种新的 QoS 参数, 下面将详细介绍本申请实施例提供的基于数据包组的 QoS 参数和混合 QoS 参数。

基于数据包组的 QoS 参数可能包括下述中的一个或多个: 组 5G QoS 标识 (Group-5G QoS Identifier, G-5QI)、组优先级 (Group Priority Level, GPL)、组延迟预算 (Group Delay Budget, GDB)、组错误率 (Group Error Rate, GER)、聚合组错误率 (Aggregation Group Error Rate, AGER)、最大组丢失率 (Maximum Group Loss Rate, MGLR)、最大聚合组

丢失率 (Maximum Aggregation Group Loss Rate, MAGLR)、平均窗口 (Averaging Window)、最大数据突发量 (Maximum Data Burst Volume)。

应理解, 上述基于数据包组的 QoS 参数的名称仅为示例, 但并不限于此。例如, 组 5G QoS 标识还可以称为媒体单元 5G QoS 标识 (Media Unit 5G QoS Identifier, MU-5QI), 组优先级还可以称为媒体单元优先级 (Media Unit Priority Level, MUPL), 组延迟预算还可以称为媒体单元延迟预算 (Media Unit Delay Budget, MUDB)、组错误率还可以称为媒体单元错误率 (Media Unit Error Rate, MUER)、聚合组错误率还可以称为聚合媒体单元错误率 (Aggregation Media Unit Error Rate, AMUER)、最大组丢失率还可以称为最大媒体单元丢失率 (Maximum Media Unit Loss Rate, MMULR)、最大聚合组丢失率还可以称为最大聚合媒体单元丢失率 (Aggregation Maximum Media Unit Loss Rate, MAMULR)。

组优先级用于指示数据包组的优先级, 同一业务的数据流的不同数据包组可以有不同的优先级。例如, 应用服务器可以定义不同数据包组的重要性, 则不同重要性的数据包组可以对应不同的 GPL。类似地, 媒体单元优先级用于指示媒体单元的优先级, 同一业务的不同媒体单元可以有不同的优先级。

组延迟预算用于指示数据包组在终端设备和锚点 UPF 之间延迟的上限。组延迟预算可以理解为将数据组中的所有数据包作为一个整体保障时延, 即数据组中的每一个数据包的时延为 GDB 与数据包组的最后一个数据包的到达时间之和。例如, 第一数据包组的第一个数据包在 1ms 到达, 最后一个数据包在 2ms 到达, 组延迟预算为 10ms, 则第一数据包组中的所有数据包需要在 12ms 前发送到接收端。类似地, 媒体单元延迟预算用于指示媒体单元在终端设备和锚点 UPF 之间延迟的上限, 即媒体单元中的每一个数据包的时延为媒体单元延迟预算与媒体单元的最后一个数据包的到达时间之和。

组错误率用于指示数据包组传输错误的上限, 也可以理解为未成功传送的数据包组的上限。当数据包组中的一个或多个数据包出现传输错误使得该数据包组无法解码时, 可以该数据包组判定为传输错误, 计入组错误率。以组错误率为 0.02 为例, 假设第一 QoS 流传输的数据流包括 1000 个数据包组, 则需要保障该第一 QoS 流中传输错误的数据包组的个数在 20 个以下。类似地, 媒体单元错误率用于指示媒体单元传输错误的上限, 也可以理解为未成功传送的媒体单元的上限。

聚合组错误率用于指示基于重要性的数据包组传输错误的上限, 也可以理解为未成功传送的加权数据包组的上限。数据包组有着不同的重要性, 则不同重要性的数据包组可以对应不同的加权因子, 将该加权因子和传输错误的数据包组的个数一起计入计算。例如, 第一数据包组对应的加权因子为 0.5, 则当第一数据包组传输错误时, 则可以计为 0.5 个数据包组传输错误, 再例如, 第二数据包组对应的加权因子为 1, 则当第二数据包组传输错误时, 则可以记为 1 个数据包组传输错误。以 AGER 为 0.02 为例, 假设第一 QoS 流传输的数据流包括 1000 个数据包组, 其中有 15 个加权因子为 1 的数据包组传输错误, 10 个加权因子为 0.4 的数据包组传输错误, 则可以计为 19 个数据包组传输错误, 则该第一 QoS 流满足聚合组错误率。类似地, 聚合媒体单元错误率用于指示基于重要性的媒体单元传输错误的上限, 也可以理解为未成功传送的加权媒体单元的上限。

最大组丢失率用于指示数据包组丢失的上限, 也可以理解为可容忍的丢弃的数据包组的上限。当数据包组中的一个或多个数据包出现丢失使得该数据包组无法解码时, 可以该

数据包组判定为丢失，计入最大组丢失率。以最大组丢失率为 0.02 为例，假设第一 QoS 流传输的数据流包括 1000 个数据包组，则需要保障该第一 QoS 流中丢失的数据包组的个数在 20 个以下。类似的，最大媒体单元丢失率用于指示媒体单元丢失的上限，也可以理解为可容忍的丢弃的媒体单元的上限。

- 5 在本申请的一些实施例中，当 QoS 流的类型为保证比特速率（Guaranteed Bit Rate, GBR）QoS 流时，可以使用组丢失率。

最大聚合组丢失率用于指示基于重要性的数据包组传输丢失的上限，也可以理解为可容忍的丢弃的加权数据包组的上限。数据包组有着不同的重要性，则不同重要性的数据包组可以对应不同的加权因子，将该加权因子和传输丢失的数据包组的个数一起计入计算。

- 10 例如，第一数据包组对应的加权因子为 0.5，则当第一数据包组传输丢失时，则可以计为 0.5 个数据包组传输丢失，再例如，第二数据包组对应的加权因子为 1，则当第二数据包组传输丢失时，则可以记为 1 个数据包组传输丢失。以最大聚合组丢失率为 0.02 为例，假设第一 QoS 流传输的数据流包括 1000 个数据包组，其中有 15 个加权因子为 1 的数据包组传输丢失，10 个加权因子为 0.4 的数据包组传输丢失，则可以计为 19 个数据包组传输丢失，则该第一 QoS 流满足聚合最大组丢失率。类似地，最大聚合媒体单元丢失率用于指示基于重要性的媒体单元传输丢失的上限，也可以理解为可容忍的丢弃的加权媒体单元的上限。

应理解，针对媒体单元的描述可以参见针对数据包组的描述，为了简洁，在此不再赘述。

- 20 G-5QI 是一个索引值，用于关联一个或多个基于数据包组的 QoS 参数。示例性地，表 1 所示为基于数据包组的 QoS 参数。如表 1 所示，当 G-5QI=100 时，QoS 流的类型为 Non-GBR、组优先级为 68、组延迟预算为 20ms、聚合组错误率或组错误率为 0.01。当 G-5QI=101 时，QoS 流的类型为 GBR、组优先级为 25、组延迟预算为 20ms、聚合组错误率或组错误率为 0.01。

- 25 在一些实施例中，在可以确定数据包组的加权因子的情况下，可以使用聚合组错误率，在不能确定数据包组的加权因子的情况下，可以使用组错误率。类似地，在可以确定数据包组的加权因子的情况下，可以使用聚合最大组丢失率，在不能确定数据包组的加权因子的情况下，可以使用最大组丢失率。

表 1 基于数据包组的 QoS 参数。

组 5G QoS 标识	资源类型	组优先级	组延迟 预算	聚合组错误 率/组错误率	最大数据 突发量	平均窗口	服务示例
100	Non-GBR	68	20ms	10^{-2}	N/A	N/A	交互服务
101	GBR	25	20ms	10^{-2}	63000 bytes	2000ms	交互服务

- 30 替代性的，本申请还提出了一种混合 QoS 参数，混合 QoS 参数是指包括基于数据包组的 QoS 参数和基于数据包组的 QoS 参数。混合 QoS 参数可能包括下述中的一个或多个：5G QoS 标识（5G QoS Identifier, 5QI）、分配保持优先级（allocation and retention priority, ARP）、反射 QoS 属性（Reflective QoS Attribute, RQA）、通知控制（Notification control）、流比特率（Flow Bit Rates）、聚合比特率（Aggregate Bit Rates）、默认值（Default values）、

最大丢失率 (Maximum Loss Rate, MLR)、资源类型 (Resource Type)、优先级 (Priority Level, PL)、延迟预算 (Delay Budget, DB)、错误率 (Error Rate, ER)、平均窗口 (Averaging Window)、最大数据突发量 (Maximum Data Burst Volume) 等。

一种可能的实现方式, 可以通过 5QI 的取值判断 QoS 参数的类型。表 2 所示为一种混合 QoS 参数的形式。例如, 5QI 取值为 1-100 时, 5QI 关联的 QoS 参数为基于数据包的 QoS 参数。例如, 5QI 取值为 1 时, 错误率实质上等同于包错误率。5QI 取值为 200-300 时, 5QI 关联的 QoS 参数为基于数据包组的 QoS 参数。例如, 5QI 取值为 200 时, 错误率实质上等同于组错误率。换句话说, 可以通过 5QI 的取值指示 QoS 流的控制方式。例如, 当网络设备或终端设备接收到 5QI 的值为 1 时, 网络设备或终端设备可以以数据包为粒度控制 QoS 流, 该 QoS 流的 QoS 需求为 5QI=1 对应的 QoS 参数; 当网络设备或终端设备接收到 5QI 的值为 200 时, 接入网设备可以以数据包组为粒度控制 QoS 流, 该 QoS 流的 QoS 需求为 5QI=200 对应的 QoS 参数。

可以理解的是, 当请求信息请求 5QI=200 的 QoS 参数时, 可以理解为该请求信息请求的是 QoS 控制方式为第一控制方式。

需要说明的是, 上述 5QI 取值为 1-100 时, 5QI 关联的 QoS 参数为基于数据包的 QoS 参数, 5QI 取值为 200-300 时, 5QI 关联的 QoS 参数为基于数据包组的 QoS 参数仅为示例, 本申请实施例对此并不限定。

表 2 一种混合 QoS 参数

5G QoS 标识	资源类型	优先级	延迟预算	错误率	最大数据突发量	平均窗口	服务示例
1	Non-GBR	68	10ms	0.0001	N/A	N/A	交互服务
....							
100	GBR	25	10ms	0.0001	N/A	N/A	交互服务
....							
200	Non-GBR	68	15ms	0.01	N/A	N/A	交互服务
...							
300	GBR	25	15ms	0.01	63000 bytes	2000ms	交互服务

一种可能的实现方式, 混合 QoS 参数还可以包括控制模式 (control mode)。控制模式用于指示 QoS 流控制方式。表 3 所示为一种混合 QoS 参数的形式。例如, 5QI 取值为 1 时, control mode 为 packet, 即表示 QoS 流的控制方式为基于数据包控制, 则对应的 QoS 参数为基于数据包的 QoS 参数。再例如, 5QI 取值为 200 时, control mode 为 group (或 media unit), 即表示 QoS 流的控制方式为基于数据包组控制, 则对应的 QoS 参数为基于数据包组的 QoS 参数。换句话说, 可以通过 5QI 指示 QoS 流的控制方式。例如, 当网络设备或终端设备接收到 5QI 的值为 1 时, 网络设备或终端设备可以以数据包为粒度控制 QoS 流; 当网络设备或终端设备接收到 5QI 的值为 200 时, 网络设备或终端设备可以以数据包组为粒度控制 QoS 流。

可以理解的是, 当请求信息请求 QoS 参数对应的控制模式为 group 时, 可以理解为该请求信息请求的是 QoS 控制方式为第一控制方式。

表 3 一种混合 QoS 参数

5G QoS 标识	控制模式	资源类型	优先级	延迟预算	错误率	最大数据突发量	平均窗口	服务示例
1	Packet	Non-GBR	68	10ms	0.0001	N/A	N/A	交互服务
....								
100	Packet	GBR	25	10ms	0.0001	N/A	N/A	交互服务
....								
200	Group	Non-GBR	68	15ms	0.01	N/A	N/A	交互服务
...								
300	Group	GBR	25	15ms	0.01	63000 bytes	2000ms	交互服务

一种可能的实现方式，可以通过指示信息指示混合 QoS 参数的类型。表 4 所示为一种混合 QoS 参数的形式。例如，以 5QI=1 为例，当指示信息指示 QoS 流的控制方式为基于数据包控制，则 5QI=1 对应的 QoS 参数可以理解为基于数据包的 QoS 参数，当指示信息指示 QoS 流的控制方式为基于数据包组控制，则 5QI=1 对应的 QoS 参数可以理解为基于数据包组的 QoS 参数。

表 4 一种混合 QoS 参数

5G QoS 标识	资源类型	优先级	延迟预算	错误率	最大数据突发量	平均窗口	服务示例
1	Non-GBR	68	10ms	0.0001	N/A	N/A	交互服务
....							
100	GBR	25	10ms	0.0001	N/A	N/A	交互服务
....							
200	Non-GBR	68	15ms	0.01	N/A	N/A	交互服务
...							
300	GBR	25	15ms	0.01	63000 bytes	2000ms	交互服务

需要说明的是，本申请实施例中，可以以混合 QoS 参数替代基于数据包组的 QoS 参数和基于数据包的 QoS 参数，因此混合 QoS 参数也可以称为 QoS 参数。

在另一些实施例中，该请求信息可以直接请求 QoS 控制方式。示例性的，该请求信息中包括第一参数，该第一参数可以为 1 个比特或几个比特，该第一参数用于指示请求的 QoS 控制方式。例如，该请求信息中包括第一参数，该第一参数为 1 个比特，其中 0 表示请求第一控制方式，1 表示请求第二控制方式。

S402，第一网元发送第二信息，该第二信息用于指示第一 QoS 流的控制方式和 QoS 参数，该第一 QoS 流用于传输该第一业务的数据流，该 QoS 控制方式包括基于数据包组控制的第一控制方式。

具体地，第一网元接收第一信息后，可以根据第一信息请求确定第一 QoS 流的控制

方式和 QoS 参数, 然后向第一设备、接入网设备、第二网元发送一个或多个参数 (即第二信息), 以使第一设备、接入网设备、第二网元根据第二信息对第一业务的数据流执行 QoS 控制。第二信息可以是一个参数, 该参数即为第一 QoS 流的 QoS 参数; 第二信息还可以是多个参数, 例如 2 个参数, 其中一个参数用于指示第一 QoS 流的控制方式, 另一个参数为第一 QoS 流的 QoS 参数。

可选的, 在一些实施例中, 第二信息包括第一 QoS 参数, 该第一 QoS 参数包括 N 个基于数据包组的 QoS 参数, $N \geq 1$ 且 N 为正整数。可以理解的是, 第一网元可以通过该第一 QoS 参数指示第一 QoS 流的控制方式。例如, 第一 QoS 参数为 G-5QI 时指示了第一 QoS 流的控制方式为第一控制方式。

示例性的, 第一设备为第一业务请求 QoS, 第一网元接收第一信息后根据第一信息确定第一 QoS 流的控制方式为第一控制方式和对应的基于数据包组的第一 QoS 参数, 然后向第一设备、接入网设备、第二网元发送第一 QoS 参数, 从而指示了第一 QoS 流的控制方式以及对应的 QoS 参数。例如, 第一网元向第一设备、接入网设备、第二网元发送 G-5QI, 从而第一设备、接入网设备、第二网元可以根据该 G-5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。再例如, 第一网元向第一设备、接入网设备、第二网元发送取值为 200 的 5QI, 该 5QI 对应的是基于数据包组的 QoS 参数 (例如通过协议事先约定 5QI 为 200 时, 对应的 QoS 参数为基于数据包组的 QoS 参数), 从而第一设备、接入网设备、第二网元可以根据该 5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。再例如, 第一网元向第一设备、接入网设备、第二网元发送取值为 200 的 5QI, 该 5QI 对应的控制模式为 group, 从而第一设备、接入网设备、第二网元可以根据该 5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。

示例性的, 第一设备请求基于数据包组控制的第一控制方式, 第一网元接收第一信息后根据第一信息确定第一 QoS 流的控制方式为第一控制方式和对应的基于数据包组的第一 QoS 参数, 然后向第一设备、接入网设备、第二网元发送第一 QoS 参数, 从而指示了第一 QoS 流的控制方式以及对应的 QoS 参数。例如, 第一网元向第一设备、接入网设备、第二网元发送 G-5QI, 从而第一设备、接入网设备、第二网元可以根据该 G-5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。再例如, 第一网元向第一设备、接入网设备、第二网元发送取值为 200 的 5QI, 该 5QI 对应的是基于数据包组的 QoS 参数 (例如通过协议事先约定 5QI 为 200 时, 对应的 QoS 参数为基于数据包组的 QoS 参数), 从而第一设备、接入网设备、第二网元可以根据该 5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。再例如, 第一网元向第一设备、接入网设备、第二网元发送取值为 200 的 5QI, 该 5QI 对应的控制模式为 group, 从而第一设备、接入网设备、第二网元可以根据该 5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。

示例性的, 第一设备请求基于数据包控制的第二控制方式, 第一网元接收第一信息后可以根据第一业务的特点确定第一 QoS 流的控制方式为基于数据包组控制的第一控制方式和对应的基于数据包组的第一 QoS 参数, 例如, 第一网元确定第一业务为 XR 等媒体类业务时, 则可以确定第一 QoS 流的控制方法为第一控制方式。换句话说, 第一网元可以根据业务的特点最终确定第一 QoS 流的控制方式, 然后通过第一 QoS 参数指示第一 QoS 流的控制方式。例如, 第一网元向第一设备、接入网设备、第二网元发送 G-5QI, 从而第

一设备、接入网设备、第二网元可以根据该 G-5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。再例如，第一网元向第一设备、接入网设备、第二网元发送取值为 200 的 5QI，该 5QI 对应的是基于数据包组的 QoS 参数（例如，通过协议事先约定 5QI 为 200 时，对应的 QoS 参数为基于数据包组的 QoS 参数），从而第一设备、接入网设备、第二网元可以根据该 5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。再例如，第一网元向第一设备、接入网设备、第二网元发送取值为 200 的 5QI，该 5QI 对应的控制模式为 group，从而第一设备、接入网设备、第二网元可以根据该 5QI 对第一业务的数据流执行以数据包组为粒度的 QoS 控制。

可选的，在一些实施例中，第二信息包括第一指示信息和第二 QoS 参数，第一指示信息用于指示第一 QoS 流的 QoS 控制方式，第二 QoS 参数用于指示第一 QoS 流的 QoS 参数。

示例性的，第一设备请求基于数据包组控制的第一控制方式或基于数据包控制的第二控制方式，第一网元接收第一信息后根据第一信息确定第一 QoS 流的 QoS 控制方式为第一控制方式和对应的 QoS 参数，然后向第一设备、接入网设备、第二网元发送第一指示信息和第二 QoS 参数，从而第一设备、接入网设备、第二网元可以根据该第一指示信息对第一业务的数据流执行以数据包组为粒度的 QoS 控制，以满足第二 QoS 参数。例如，第二 QoS 参数为表 4 中的任意一行，当第一指示信息指示基于数据包组控制时，第二 QoS 参数可以理解为基于数据包组的 QoS 参数。再例如，第二 QoS 参数为表 4 中的任意一行，当第一指示信息指示基于数据包控制时，第二 QoS 参数可以理解为基于数据包的 QoS 参数。再例如，第二 QoS 参数为基于数据包的 QoS 参数，但第二 QoS 参数是适用于第一控制方式的基于数据包的 QoS 参数，其中第一网元可以根据第一业务的数据流中数据包和数据包组的关系和/或数据包组的加权因子确定第二 QoS 参数。例如当以数据包为粒度对第一 QoS 流执行 QoS 控制时 QoS 参数为 5QI=5，第一网元可以根据数据包和数据包组的换算关系确定 5QI=100 的 QoS 参数适用于以数据包组为粒度对第一 QoS 流执行 QoS 控制，则该 5QI=100 的 QoS 参数为第二 QoS 参数。

需要说明的是，第一网元可以根据以下几种方式确定第一业务的数据流的数据包与数据包组的关系和/或数据包组的加权因子。

方式一：第一业务的服务器可以直接指示第一业务的数据流的数据包与数据包组的关系和/或加权因子。例如，第一业务的服务器可以指示第一业务的数据流包括 5 个数据包，该 5 个数据包组包括 500 个数据包，该 5 个数据包组对应的加权因子为 0.5。

方式二：第一网元可以通过第一业务的业务信息确定第一业务的数据流的数据包与数据包组的关系和/或加权因子。其中，业务信息包括第一业务的编码参数和/或第一业务中每一个数据包组的加权因子。第一业务的编码参数包括帧率、分辨率、码率等。第一网元可以根据编码参数计算得到数据包与数据包组的关系。例如，假设第一业务的每一帧的大小相同，每一帧为一个数据包组，一个数据包的大小为 10kb，第一业务的每秒传输帧数（framespersecond，FPS）为 10，码率为 10Mbps，则可以得到第一业务每一帧的大小为 1mb，则每一帧包含的数据包的个数为 100 个。

进一步的，编码参数还可以包括视频图像组（Group of picture，GoP），从而可以计算出第一业务的关键帧（I 帧）和辅助帧（P 帧）的个数，进而可以将关键帧和辅助帧对

应不同的加权因子。例如，关键帧的加权因子为 1，辅助帧的加权因子为 0.3。

针对方式二，方法 400 还包括：第一网元获取第一业务的业务信息。第一网元可以从第一设备或其他网元获取第一业务的业务信息，本申请实施例对比不作限定。

5 可选的，在一些实施例中，第二信息包括第一指示信息和第二 QoS 参数，第一指示信息用于指示第一 QoS 流的 QoS 控制方式，第二 QoS 参数包括至少一个基于数据包组的 QoS 参数。

示例性的，第一设备请求基于数据包组控制的第一控制方式或基于数据包控制的第二控制方式，第一网元接收第一信息后根据第一信息确定第一 QoS 流的 QoS 控制方式为第一控制方式和第二 QoS 参数，其中，第二 QoS 参数为基于数据包组的 QoS 参数。需要说明的是，第一网元确定的第二 QoS 参数是能够满足第一业务的数据流以数据包为粒度的 QoS 需求。第一设备、第二网元、接入网设备接收第一指示信息和第二 QoS 参数后，可以根据第一业务的数据流中数据包与数据包组的关系最终确定能够满足以数据包组为粒度执行 QoS 控制的 QoS 参数。例如，第一网元确定的第二 QoS 参数中 5QI 对应的包错误率为 0.0001，第一设备、第二网元、接入网设备接收第一指示信息和第二 QoS 参数后，确定第一业务的数据流包括 10000 数据包，该 10000 个数据包可以组成 100 个数据包组，则第一设备、第二网元、接入网设备可以最终确定组错误率为 0.01。换句话说，第一网元可以先确定满足以数据包为粒度执行 QoS 控制的 QoS 参数，由第一设备、第二网元、接入网设备最终确定满足以数据包组为粒度执行 QoS 控制的 QoS 参数。

针对该实施例，方法 400 还包括：第一网元发送第一业务的业务信息。
20 可选的，在一些实施例中，第一指示信息具体用于指示第一 QoS 流的优选控制方式，也可以称为控制方式优先级。

第一网元可以通过第一指示信息指示第一 QoS 流的优选控制方式，由第一设备、第二网元、接入网设备根据第一指示信息最终确定对应的 QoS 参数，具体描述请参加后文描述。

25 可选的，在一些实施例中，第二信息包括第二指示信息、第三 QoS 参数和第四 QoS 参数，其中第二指示信息用于指示第一 QoS 流的优选 QoS 控制方式，第三 QoS 参数包括 M 个基于数据包组的 QoS 参数， $M \geq 1$ 且 M 为正整数，第四 QoS 参数包括 L 个基于数据包组的 QoS 参数， $L \geq 1$ 且 L 为正整数。

第一网元可以通过第二指示信息指示第一 QoS 流的优选 QoS 控制方式，由第一设备、第二网元、接入网设备根据第二指示信息最终确定采用第三 QoS 参数还是第四 QoS 参数。

应理解，第二指示信息通过指示第一 QoS 流的优选 QoS 控制方式间接指示了第一 QoS 流的 QoS 控制方式，也即第二信息用于指示第一 QoS 流的控制方式。

S403，第一设备根据第二信息对该第一业务的数据流执行 QoS 控制。

具体的，第一设备接收第二信息后可以根据第二信息对该第一业务的数据流执行 QoS 控制。第一设备可以基于数据包组为粒度对该第一业务的数据流执行 QoS 控制。例如，当第一业务的数据流中的第一数据包组中的一个数据包发生丢失时，第一设备可以丢弃第一数据包组中的其他数据包。再例如，第一设备可以优先保障传输比例高的数据包组。再例如，第一设备可以优先保障重要性高的数据包组。

可选的，在一些实施例中，第二信息包括第一 QoS 参数，该第一 QoS 参数包括 N 个

基于数据包组的 QoS 参数, $N \geq 1$ 且 N 为正整数。

第一设备接收第一 QoS 参数后可以确定第一 QoS 流的控制方式为以数据包组控制的第一控制方式, 并基于第一 QoS 参数对第一业务的数据流执行 QoS 控制。针对第一 QoS 参数指示第一 QoS 流的控制方式的描述可以参见前文描述, 在此不再赘述。

- 5 可选的, 在一些实施例中, 第二信息包括第一指示信息和第二 QoS 参数, 第一指示信息用于指示第一 QoS 流的 QoS 控制方式, 第二 QoS 参数用于指示第一 QoS 流的 QoS 参数。

- 10 第一设备接收第一指示信息和第二 QoS 参数后, 可以根据第一指示信息指示的控制方式对该第一业务的数据流执行 QoS 控制以满足第二 QoS 参数。针对第二 QoS 参数和第一指示信息的描述可以参见前文描述, 在此不再赘述。

可选的, 在一些实施例中, 第二信息包括第一指示信息和第二 QoS 参数, 第一指示信息用于指示第一 QoS 流的 QoS 控制方式, 第二 QoS 参数包括至少一个基于数据包组的 QoS 参数。

- 15 第一设备接收第一指示信息和第二 QoS 参数后, 可以根据第一指示信息指示的 QoS 控制方式执行 QoS 控制。例如, 当第一指示信息指示基于数据包控制的第二控制方式时, 第一设备可以以数据包为粒度对该第一业务的数据流执行 QoS 控制以满足第二 QoS 参数; 当第一指示信息指示基于数据包组控制的第一控制方式时, 第一设备可以根据第一业务的数据流中数据包和数据包组的关系和/或数据包组的加权因子确定适用于以数据包组为粒度控制的 QoS 参数, 然后第一设备以数据包组为粒度对该第一业务的数据流执行 QoS 控制以满足确定的 QoS 参数。

需要说明的是, 第一设备可以根据以下几种方式确定第一业务的数据流的数据包与数据包组的关系和/或数据包组的加权因子。

方式一: 第一业务的服务器可以直接指示第一业务的数据流的数据包与数据包组的关系和/或加权因子。

- 25 方式二: 第一设备可以通过第一业务的业务信息确定第一业务的数据流的数据包与数据包组的关系和/或加权因子。

针对方式二, 方法 400 还包括: 第一设备获取第一业务的业务信息。第一设备可以从核心网网元获取第一业务的业务信息。

- 30 可选的, 在一些实施例中, 第一指示信息具体用于指示第一 QoS 流的优选控制方式, 也可以称为控制方式优先级。

- 35 示例性的, 第一设备接收第一指示信息后, 当第一设备支持第一指示信息指示的优选 QoS 控制方式时, 第一设备执行第一指示信息指示的 QoS 控制方式, 当第一设备不支持第一指示信息指示的 QoS 控制方式时, 执行另一种控制方式。例如, 第一指示信息指示的优选 QoS 控制方式为第一控制方式, 第二 QoS 参数为基于数据包组的 QoS 参数, 且第一设备支持第一控制方式, 则第一设备根据第一业务的数据流的数据包与数据包组的关系和/或数据包组的加权因子确定满足第一控制方式的 QoS 参数; 当第一设备不支持第一控制方式时, 第一设备执行第二控制方式, 第二 QoS 参数即为满足 QoS 需求的 QoS 参数。

通过上述技术方案, 第一设备可以灵活的选择适合的 QoS 控制方式, 提高了 QoS 管理的灵活性。

可选的, 在一些实施例中, 第二信息包括第二指示信息、第三 QoS 参数和第四 QoS 参数, 第二指示信息用于指示第一 QoS 流的优选 QoS 控制方式, 第三 QoS 参数包括 M 个基于数据包组的 QoS 参数, $M \geq 1$ 且 M 为正整数, 第四 QoS 参数包括 L 个基于数据包的 QoS 参数, $L \geq 1$ 且 L 为正整数。

- 5 第一设备接收第二指示信息后, 当第一设备支持第二指示信息指示的优选 QoS 控制方式时, 第一设备执行第二指示信息指示的 QoS 优选控制方式, 当第一设备不支持第二指示信息指示的 QoS 优选控制方式时, 执行另一种控制方式。例如, 第二指示信息指示的优选 QoS 控制方式为第一控制方式, 且第一设备支持第一控制方式, 则第一设备执行第一控制方式, 第三 QoS 参数即为满足 QoS 需求的 QoS 参数; 当第一设备不支持第一控制方式时, 则第一设备执行第二控制方式, 第四 QoS 参数即为满足 QoS 需求的 QoS 参数。

10 通过上述技术方案, 第一设备可以灵活的选择适合的 QoS 控制方式, 提高了 QoS 管理的灵活性。

S404, 接入网设备根据第二信息对该第一业务的数据流执行 QoS 控制。

S405, 第二网元根据第二信息对该第一业务的数据流执行 QoS 控制。

- 15 可以理解的是, 针对接入网设备和第二网元根据第二信息对该第一业务的数据流执行 QoS 控制的描述可以参见 S403, 在此不再赘述。

可选的, 在第一网元发送第二信息之前, 方法 400 还包括: 第一网元根据所述第一业务确定所述第一 QoS 流的 QoS 控制方式, 或第一网元根据所述第一设备是否支持第一控制方式确定所述第一 QoS 流的 QoS 控制方式。

- 20 第一网元可以根据第一业务的特点或第一设备是否支持第一控制方式确定第一 QoS 流的控制方式。例如, 当第一设备不支持第一控制方式时, 则第一网元不会指示第一设备执行第一控制方式。

- 25 本申请实施的 QoS 管理方法, 第一网元接收请求 QoS 的信息后, 可以根据第一业务的特点确定不同的 QoS 控制方式, 可以为第一业务提供基于数据包组的 QoS 控制方式以及对应的 QoS 参数, 提高了 QoS 管理的灵活性, 满足了不同业务的需求, 提高了用户体验, 降低了网络资源的浪费。

- 30 应理解, 在上述步骤 S401-S405 中, (1) 第一网元可以为 SMF; (2) 第二网元可以为 UPF; (3) 第一设备可以是终端设备或应用服务器。为了便于理解本申请提供的一种 QoS 管理的方法, 以下, 作为示例而非限定, 以第一网元为 SMF, 第二网元为 UPF, 第一设备为终端设备为例, 参考如图 5 所示的具体示例方法, 分别对方法 400 中的 S401-S405 进行详细说明。

需要说明的是, 下面提到的某些步骤与上述方法 400 相同, 此处对于相关细节不再赘述, 具体过程可参考方法 400 相关步骤。方法 500 以下行数据为例进行说明。

- 35 S501, 终端设备向 SMF 发送第一信息, 该第一信息用于为第一业务请求 QoS。

在一种可能的实现方式, 终端设备通过 AMF 向 SMF 发送第一信息。

应理解, 终端设备向 SMF 发送第一信息之前, 终端设备已经通过 SMF 建立 PDU 会话, 并通过 PDU 会话与应用服务器建立连接, 则第一信息可以包括在对 PDU 会话请求修改的会话修改请求信息中。

S502, SMF 发送第二信息, 该第二信息用于指示第一 QoS 流的控制方式和 QoS 参数,

该第一 QoS 流用于传输该第一业务的数据流，该 QoS 控制方式包括基于数据包组控制的第一控制方式。

具体的，SMF 向终端设备、接入网设备、UPF 发送第二信息。

- 一种可能的实现方式中，SMF 通过 AMF 向接入网设备发送第二信息，通过 AMF 和接入网设备向终端设备发送第二信息。第二信息可以包括在 QoS 配置 (QoS profile) 中向接入网设备发送，第二信息可以包括在 QoS 规则 (QoS rule) 中向终端设备发送，第二信息可以包括在包检测规则 (Packet detection rule) 中向 UPF 设备发送。

S503，应用服务器向 UPF 发送第一业务的数据流。

- 可选的，在一些实施例中，应用服务器可以在 S503 中指示第一业务的数据流的数据包和数据包组的关系。例如，第一业务的数据流包括 100 个数据包，该 100 个数据包可以组成 3 个数据包组。

一种可能的实现方式，应用服务器在第一业务的数据包的包头增加指示信息。

- 示例性的，第一业务的每一个数据包的包头中包括指示信息，该指示信息用于指示该数据包属于哪一个数据包组 (即所属数据包组的序号)、该数据包在所属数据包组中的位置 (即数据包组中具体的包序号)、该数据包所属数据包组开始的标志 (即数据包组中的第一个数据包) 和结束的标志 (即数据包组中的最后一个数据包)、该数据包所属数据包组的重要性中的至少一项。例如，第一数据包的包头的指示信息用于指示第一数据包属于第一数据包组，第一数据包位于第一数据包组的第一位置，第一数据包组起始于第一数据包，结束于第四数据包中的至少一项。

- 示例性的，第一业务的每一个数据包组的第一个数据包的包头中包括指示信息，该指示信息用于指示该数据包属于哪一个数据包组 (即所属数据包组的序号)、该数据包所属数据包组的大小 (即所述数据包组共有多少个数据包)、该数据包所属数据包组开始的标志 (即数据包组中的第一个数据包)、该数据包所属数据包组的上一个数据包组结束的标志、该数据包所属数据包组的重要性中的至少一项。例如，第一数据包组包括第一数据包、第二数据包、第三数据包且第一数据包为该第一数据包组的第一个数据包，则该第一数据包的包头中包括指示信息，该指示信息指示第一数据包属于第一数据包组，第一数据包组的大小为 3 (第二数据包和第三数据包与第一数据包同属于第一数据包组)，第一数据包位于第一数据包组的第一位置，第一数据包组起始于第一数据包，结束于第三数据包中的至少一项。

- 示例性的，第一业务的每一个数据包组的第一个数据包和最后一个数据包的包头中包括指示信息，该指示信息用于指示该数据包属于哪一个数据包组 (即所属数据包组的序号)、该数据包所属数据包组的大小 (即所述数据包组共有多少个数据包)、第一个数据包的包头指示信息指示该数据包为所属数据包组开始的标志、最后一个数据包的包头指示信息指示该数据包为所属数据包组结束的标志。例如，第一数据包组包括第一数据包、第二数据包、第三数据包且第一数据包为该第一数据包组的第一个数据包，第三数据包为该第一数据包组的最后一个数据包，则第一数据包和第三数据包的包头中包括指示信息，该指示信息指示第一数据包及第三数据包属于第一数据包组，第一数据包组的大小为 3 (第二数据包与第一数据包与第三数据包同属于第一数据包组)，第一数据包的包头的指示信息指示第一数据包为第一数据包组的起始数据包，第三数据包的包头的指示信息指示第三数据包

为第一数据包组的结束数据包中的至少一项。

示例性的，第一业务的第一个数据包头包括指示信息，该指示信息用于指示该第一业务的所有数据包所属的数据包组、每一个数据包组开始的标识和结束的标识、每一个数据包在所属数据包组中的位置中的至少一项。例如，第一数据包为该第一业务的第一个数据包，该第一数据的包头包括指示信息，该指示信息指示第一数据包、第二数据包、第三数据包为属于第一数据包组、第一数据包组起始于第一数据包，结束于第三数据包，第一数据包位于第一数据包组的第一位置，第二数据包位于第一数据包组的第二位置，第三数据包位于第一数据包组的第三位置。

10 可选的，在一些实施例中，应用服务器还可以在 S503 中指示第一业务的数据流中每一个数据包组的加权因子。加权因子可以对应不同的重要性，即可以理解为对于用户影响的程度。以 XR 业务为例，XR 业务的关键帧对于用户的影响程度大，辅助帧对应用户的影响程度小，因此 XR 业务的关键帧的加权因子大于辅助帧的加权因子。

针对应用服务器指示每一个数据包组的加权因子的描述类似于应用服务器指示第一业务的数据流的数据包和数据包组的关系的描述，为了简洁，在此不再赘述。

15 一种可能的实现方式，应用服务器可以单独发送一个指示信息用于指示第一业务的数据包组与数据包的关系和/或每一个数据包组的加权因子。

S504，UPF 向接入网设备发送第一业务的数据流。

具体的，UPF 可以向接入网设备发送第一业务的数据，并根据第二信息对第一业务的数据流执行 QoS 控制。

20 可选的，在一些实施例中，UPF 可以在第一业务的数据包的包头中增加标记信息，例如添加在 GTP-U 包头中，用于指示第一业务的数据包组的数量和/或每一个数据包组的起始数据包和终止数据包和/或每一个数据包的加权因子和/或每一个数据包组包括的数据包的数量。

25 一种可能的实现方式，UPF 可以通过应用服务器指示得到该标记信息，即通过 S503 步骤获取。

一种可能的实现方式，UPF 可以根据第一业务的业务信息得到该标记信息。

S505，接入网设备向终端设备发送第一业务的数据流。

30 具体的，接入网设备接收到 UPF 发送的第一业务的数据流后，可以根据第二信息对第一业务的数据流执行 QoS 控制。例如，接入网设备可以基于组延迟预算调度第一业务的数据包，保证同一个数据包组的数据包的延迟符合组延迟预算。再例如，接入网设备可以基于组错误率调度第一业务的数据包，当一个数据包组中的一个数据包传输错误时，接入网设备丢弃该数据包组的其他数据包。再例如，接入网设备可以优先保障传输完成比例大的数据包组，以免造成已经传输的数据包的无效传输。

S506，终端设备根据第二信息对第一业务的数据流执行 QoS 控制。

35 具体的，终端设备可以根据第二信息对第一业务的数据流执行 QoS 控制。例如，终端设备可以基于组错误率对第一业务的数据流执行 QoS 控制，当一个数据包组的一个数据包传输错误或终端设备解码失败时，终端设备可以不再接收该数据包组的其他数据包，直至终端设备接收到传输失败或解码失败的数据包的重传数据包时，终端设备再接收该数据包组的其他数据包。

应理解,方法 500 中以传输下行数据为例,但本申请实施例并不限于此,本申请实施例还可以用于传输上行数据。例如,终端设备向接入网设备发送第一业务的数据,并根据第二信息对第一业务的数据流执行 QoS 控制。

5 应理解,当传输上行数据时,终端设备可以类似于应用服务器指示第一业务的数据包与数据包组的关系和/或每一个数据包组的加权因子。

还应理解,针对终端设备发送第一业务的数据,并根据第二信息对第一业务的数据流执行 QoS 控制的描述,可以参见前文,在此不再赘述。

本申请实施的 QoS 管理方法,第一网元接收请求 QoS 的信息后,可以根据第一业务的特点确定不同的 QoS 控制方式,可以为第一业务提供基于数据包组的 QoS 控制方式以及对应的 QoS 参数,提高了 QoS 管理的灵活性,满足了不同业务的需求,提高了用户体验,降低了网络资源的浪费。

以下,作为示例而非限定,以第一网元为 SMF,第二网元为 UPF,第一设备为应用服务器为例,参考如图 6 所示的具体示例方法,分别对方法 400 中的 S401-S405 进行详细说明。

15 S601,应用服务器向 SMF 发送第一信息,该第一信息用于为第一业务请求 QoS。

在一种可能的实现方式,应用服务器通过 PCF 向 SMF 发送第一信息。

在一种可能的实现方式中,应用服务器通过 NEF 和 PCF 向 SMF 发送第一信息。

应理解,应用服务器向 SMF 发送第一信息之前,应用服务器已经通过 PDU 会话与应用服务器建立连接,则第一信息可以包括在会话修改请求信息中。

20 S602,SMF 发送第二信息,该第二信息用于指示第一 QoS 流的控制方式和 QoS 参数,该第一 QoS 流用于传输该第一业务的数据流,该 QoS 控制方式包括基于数据包组控制的第一控制方式。

S603,应用服务器向 UPF 发送第一业务的数据流。

S604,UPF 向接入网设备发送第一业务的数据流。

25 S605,接入网设备向终端设备发送第一业务的数据流。

S606,终端设备根据第二信息对第一业务的数据流执行 QoS 控制。

应理解,针对步骤 S602-S606 的描述,可以参见前文描述,为了简洁,在此不再赘述。

本申请实施的 QoS 管理方法,第一网元接收请求 QoS 的信息后,可以根据第一业务的特点确定不同的 QoS 控制方式,可以为第一业务提供基于数据包组的 QoS 控制方式以及对应的 QoS 参数,提高了 QoS 管理的灵活性,满足了不同业务的需求,提高了用户体验,降低了网络资源的浪费。

需要说明的是,上述方法 400-600 中,终端设备发送第一信息之前已经与应用服务器建立连接,或应用服务器发送第一信息之前已经与终端设备建立连接,但本申请并不限于此,例如,终端设备与应用服务器建立连接时可以发送第一信息,从而当终端设备与应用服务器建立连接后,可以执行相应的 QoS 控制方式。

上述方法以第一业务的数据流由一个 QoS 流承载为例,但本申请并不限于此,例如,第一业务的数据流可以由两个 QoS 流承载,并可以对该两个 QoS 流执行联合控制。下面将详细介绍对多个 QoS 流执行联合 QoS 控制的 QoS 管理方法。

图 7 所示为本申请提供的一种 QoS 管理方法 700 的示意性流程图。

S701, 第一设备发送第三信息, 该第三信息用于为第二业务请求 QoS。

具体地, 第一设备向第一网元发送第三信息, 该第三信息用于为第二业务请求 QoS。

一种可能的实现方式, 该第一信息包括第一业务的标识 (例如, 应用程序标识等)、五元组 (源 IP 地址、源端口号、目标 IP 地址、目标端口号、传输层协议)、三元组 (目标 IP 地址、目标端口号、传输层协议) 中的至少一个以及请求信息, 该请求信息用于为第二业务请求 QoS。

可选的, 在另一种可能的实现方式中, 该请求信息还用于请求 QoS 流的控制方式。例如, 该请求信息可以用于请求对第二业务的数据流执行联合 QoS 控制。联合 QoS 控制可以理解对多个具有关联关系的 QoS 流执行 QoS 控制以使该多个 QoS 流共同满足对应的 QoS 参数。假设第二 QoS 流和第三 QoS 流具有关联关系, 其中第二 QoS 流传输的数据流包括 1000 个数据包, 第二 QoS 流传输的数据流包括 1000 个数据包, 如果对第二 QoS 流和第三 QoS 流执行联合 QoS 控制以使丢失的数据小于 40 个, 则需要第二 QoS 流和第三 QoS 流上丢失的数据包的之和小于 40。

在一些实施例, 该请求信息请求的是联合 QoS 参数。可以理解的是, 当该请求信息请求的联合 QoS 参数, 则该请求信息请求的 QoS 控制方式为联合 QoS 控制。下面将详细介绍本申请实施例提供的联合 QoS 参数。

联合 QoS 参数是指具有关联关系的多个 QoS 流的 QoS 需求。具有关联关系的 QoS 流可以理解为 QoS 流传输的数据流具有关联关系。例如, 第一 QoS 流与第二 QoS 流具有关联关系, 其中第一 QoS 流传输第一业务的第一数据流, 第二 QoS 流传输第一业务的第二数据流。

联合 QoS 参数可能包括下述中的一个或多个: 联合 5G QoS 标识 (CorrelatedQoS flow 5G QoS Identifier, CQF-5QI)、联合错误率 (CorrelatedQoS flow Error Rate, CQF-ER)、聚合联合错误率 (CorrelatedQoS flow Aggregation Error Rate, CQF-AER)、联合最大丢失率 (CorrelatedQoS flow Maximum Loss Rate, CQF-MLR)、聚合联合最大丢失率 (CorrelatedQoS flow Maximum Aggregation Loss Rate, CQF-MALR)、联合最大比特速率 (CorrelatedQoS flow Maximum Bit Rate, CQF-MBR)、联合聚合最大比特速率 (CorrelatedQoS flow Aggregation Maximum Bit Rate, CQF-AMBR), 下面将对联合 QoS 参数进行详细介绍。

联合错误率用于指示具有关联关系的 QoS 流中数据包组或数据包传输错误的上限, 也可以称为未成功传送的数据包组或数据包传输错误的上限。当具有关联关系的 QoS 流传输错误一个数据包, 可以将该数据包判定为传输错误, 或当具有关联关系的 QoS 流传输错误一个数据包组 (例如, 丢失该数据包组的一个数据包), 可以将该数据包组判定为传输错误。以联合错误率为 0.02 为例, 假设第一 QoS 流和第二 QoS 流具有关联关系, 其中第一 QoS 流传输的数据流包括 1000 个数据包, 第二 QoS 流传输的数据流包括 1000 个数据包, 则需要保证第一 QoS 流和第二 QoS 流中传输错误的数据包的错误个数在 40 个以下; 或第一 QoS 流传输的数据流包括 1000 个数据包组, 第二 QoS 流传输的数据流包括 1000 个数据包组, 则需要保证第一 QoS 流和第二 QoS 流中传输错误的数据包组的错误个数在 40 个以下。

聚合联合错误率用于指示基于重要性的具有关联关系的 QoS 流中数据包组或数据包

传输错误的上限,也可以理解为未成功传送的加权数据组或加权数据包传输错误的上限。数据组或数据包有着不同的重要性,则不同重要性的数据组和数据包可以对应不同的加权因子,将该加权因子和传输错误的数据组或数据包的个数一起计入计算。例如,具有关联关系的 QoS 流传输的第一数据组对应的加权因子为 0.5,则当第一数据组传输错误时,则可以计为 0.5 个数据组传输错误,再例如,第二数据组对应的加权因子为 1,则当第二数据组传输错误时,则可以记为 1 个数据组传输错误。以聚合联合错误率为 0.02 为例,假设第一 QoS 流和第二 QoS 流具有关联关系,其中第一 QoS 流传输的数据流包括 1000 个数据组,第二 QoS 流传输的数据流包括 1000 个数据组,其中第一 QoS 流有 10 个加权因子为 1 的数据组传输错误,5 个加权因子为 0.4 的数据组传输错误,第二 QoS 流有 5 个加权因子为 1 的数据组传输错误,5 个加权因子为 0.4 的数据组传输错误,则可以计为 19 个数据组传输错误,则该第一 QoS 流满足聚合联合错误率。类似地,当具有关联关系的 QoS 流传输数据包时,可以参考上述描述,在此不再赘述。

联合丢失率用于指示基于关联关系的 QoS 流中数据组或数据包丢失的上限,也可以理解为可容忍的丢弃的数据组或数据包的上限。当具有关联关系的 QoS 流丢失一个数据包,可以将该数据包判定丢失,或当具有关联关系的 QoS 流丢失一个数据组(例如,丢失该数据组的一个数据包),可以将该数据组判定丢失。以联合丢失率为 0.02 为例,假设第一 QoS 流和第二 QoS 流具有关联关系,其中第一 QoS 流传输的数据流包括 1000 个数据包,第二 QoS 流传输的数据流包括 1000 个数据包,则需要保证第一 QoS 流和第二 QoS 流中丢失的数据包的个数在 40 个以下;或第一 QoS 流传输的数据流包括 1000 个数据组,第二 QoS 流传输的数据流包括 1000 个数据组,则需要保证第一 QoS 流和第二 QoS 流中丢失的数据组的错误个数在 40 个以下。

聚合联合丢失率用于指示基于重要性的具有关联关系的 QoS 流中数据组或数据包丢失的上限,也可以理解为可容忍的丢弃的数据组或数据包的上限。数据组或数据包有着不同的重要性,则不同重要性的数据组和数据包可以对应不同的加权因子,将该加权因子和丢失的数据组或数据包的个数一起计入计算。例如,具有关联关系的 QoS 流传输的第一数据组对应的加权因子为 0.5,则当第一数据组丢失时,则可以计为 0.5 个数据组传输错误,再例如,第二数据组对应的加权因子为 1,则当第二数据组丢失时,则可以记为 1 个数据组传输错误。以聚合联合丢失率为 0.02 为例,假设第一 QoS 流和第二 QoS 流具有关联关系,其中第一 QoS 流传输的数据流包括 1000 个数据组,第二 QoS 流传输的数据流包括 1000 个数据组,其中第一 QoS 流有 10 个加权因子为 1 的数据组丢失,5 个加权因子为 0.4 的数据组丢失,第二 QoS 流有 5 个加权因子为 1 的数据组丢失,5 个加权因子为 0.4 的数据组丢失,则可以计为 19 个数据组丢失,则该第一 QoS 流满足聚合联合丢失率。类似地,当具有关联关系的 QoS 流传输数据包时,可以参考上述描述,在此不再赘述。

联合聚合最大比特速率用于指示具有一组关联关系的 QoS 流中所有类型为 Non-GBR 的 QoS 流的聚合比特速率,联合聚合最大比特速率可以通过一个平均窗口计算,即在一个平均窗口内的聚合比特速率。例如,第一 QoS 流和第二 QoS 流具有关联关系,第一 QoS 流和第二 QoS 流的类型为 Non-GBR QoS 流,其中第一 QoS 流的比特速率为 50M,第二 QoS 流的比特速率为 50M,则第一 QoS 流和第二 QoS 流的联合聚合最大比特速率为 100M。

联合最大比特速率用于指示具有一组有关联关系的 QoS 流的最大比特速率。联合最大比特速率可以通过一个平均窗口计算，即在一个平均窗口内的所有 QoS 流的聚合比特速率。例如，第一 QoS 流和第二 QoS 流具有关联关系，第一 QoS 流的类型为 GBR QoS 流，第二 QoS 流的类型为 Non-GBR QoS 流，其中第一 QoS 流的比特速率为 50M，第二 QoS 流的比特速率为 50M，则第一 QoS 流和第二 QoS 流的联合最大比特速率为 100M。

CQF-5QI 是一个索引值，用于关联一个或多个联合 QoS 参数。类似地，通过 CQF-5QI 可以得到其他联合 QoS 参数。

应理解，上述联合 QoS 参数的名称仅为示例，并不限于此。

还应理解，上述联合 QoS 参数可以分别基于数据包和基于数据包组（或媒体单元）进一步划分不同的 QoS 参数。例如，联合包最大丢失率（CorrelatedQoS flow Maximum Packet Loss Rate, CQF-MPLR）和联合媒体单元最大丢失率（CorrelatedQoS flow Media UnitMaximum Loss Rate, CQF-MMULR）分别用于指示一组具有关联关系的 QoS 流中可以丢失的数据包和媒体单元的上限；联合包错误率（CorrelatedQoS flow Packet Error Rate, CQF-PER）和联合媒体单元错误率（CorrelatedQoS flow Media Unit Error Rate, CQF-MUER）分别用于指示一组具有关联关系的 QoS 流中可以传输错误的数据包和媒体单元的上限；聚合联合包错误率（CorrelatedQoS flow PacketAggregation Error Rate, CQF-APER）和聚合联合媒体单元错误率（CorrelatedQoS flow Media UnitAggregation Error Rate, CQF-AMUER）分别用于指示一组具有关联关系的 QoS 流中可以传输错误的加权数据包和加权媒体单元的上限；聚合联合最大包丢失率（CorrelatedQoS flow MaximumAggregation Packet Loss Rate, CQF-MAPLR）和聚合联合最大媒体单元丢失率（CorrelatedQoS flowMaximumAggregation Media Unit Loss Rate, CQF-MAMULR）分别用于指示一组具有关联关系的 QoS 流中可以丢失的加权数据包和加权媒体单元的上限。

在另一些实施例中，该请求信息可以直接请求联合 QoS 控制。

替代性的，该请求信息还可以请求的是非联合 QoS 控制，但是第一网元可以根据第二业务的特点确定使用联合 QoS 控制。例如，第二业务为 XR 等媒体类业务，第二业务的 I 帧承载于第二 QoS 流，P 帧承载于第三 QoS 流，则第一网元可以确定对第二 QoS 流和第三 QoS 流执行联合 QoS 控制。

S702，第一网元发送第四信息，该第四信息包括用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数。

具体的，第一网元确定第二 QoS 流和第三 QoS 流具有关联关系，即都属于传输第二业务的 QoS 流，第一网元向第一设备、接入网设备、第二网元发送第四信息，第四信息包括用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数。

可选的，在一些实施例中，第四信息包括第五 QoS 参数，该第五 QoS 参数包括 K 个联合 QoS 参数， $K \geq 1$ 且 K 为正整数。可以理解的是，第一网元可以通过该第五 QoS 参数指示对第二 QoS 流和第三 QoS 流执行联合 QoS 控制。例如，第五 QoS 参数为 C-5QI 时指示了对第二 QoS 流和第三 QoS 流执行联合 QoS 控制。

例如，第一网元向第一设备、接入网设备、第二网元发送 C-5QI，该 C-5QI 的值为 50，从而第一设备、接入网设备、第二网元可以根据该 C-5QI 对承载第一业务的数据流的第一 QoS 流和第二 QoS 流执行联合控制。

可选的, 在一些实施例中, 所述第四信息包括第三指示信息和第六 QoS 参数, 所述第三指示信息用于指示联合控制所述第二 QoS 流和所述第三 QoS 流, 所述第六 QoS 参数为用于联合 QoS 控制的参数, 所述第六 QoS 包括 L 个基于数据包组的 QoS 参数, $L \geq 1$ 且 L 为正整数; 或所述第六 QoS 参数包括 J 个基于数据包的 QoS 参数, $J \geq 1$ 且 J 为正整数。

S703, 第一设备根据第四信息对该第二业务的数据流执行联合 QoS 控制。

具体的, 第一设备接收第四信息后可以根据第四信息对该第二业务的数据流执行 QoS 控制。例如, 第二 QoS 流承载的是加权因子=1 的数据包或数据包组, 第三 QoS 流承载的是加权因子=0.5 的数据包或数据包组, 则第一设备可以优先保障第二 QoS 流的传输。

可选的, 在一些实施例中, 第四信息包括第五 QoS 参数, 该第五 QoS 参数包括 K 个联合 QoS 参数, $K \geq 1$ 且 K 为正整数。

第一设备接收第五 QoS 参数后可以确定对第二 QoS 流和第三 QoS 流执行联合控制以满足第五 QoS 参数。

可选的, 在一些实施例中, 所述第四信息包括第三指示信息和第六 QoS 参数, 所述第三指示信息用于指示联合控制所述第二 QoS 流和所述第三 QoS 流, 所述第六 QoS 参数为用于联合 QoS 控制的参数, 所述第六 QoS 包括 L 个基于数据包组的 QoS 参数, $L \geq 1$ 且 L 为正整数; 或所述第六 QoS 参数包括 J 个基于数据包的 QoS 参数, $J \geq 1$ 且 J 为正整数。

第一设备接收第三指示信息和第六 QoS 参数后, 根据第三指示信息确定对第二 QoS 流和第三 QoS 流执行联合控制以满足第六 QoS 参数。应理解, 第六 QoS 参数可以为基于数据包组的 QoS 参数或基于数据包的 QoS 参数, 进而可以确定在执行联合控制时的粒度。例如, 第六 QoS 参数为基于数据包组的 QoS 参数, 则在第二 QoS 流的第三 QoS 流执行联合控制时是以数据包组为粒度的。

S704, 接入网设备根据第四信息对该第二业务的数据流执行联合 QoS 控制。

S705, 第二网元根据第四信息对该第二业务的数据流执行联合 QoS 控制。

可以理解的是, 针对接入网设备和第二网元根据第四信息对该第二业务的数据流执行 QoS 控制的描述可以参见 S703, 在此不再赘述。

本申请实施的 QoS 管理方法, 当业务的数据流承载于多个 QoS 流时, 第一设备可以对该多个 QoS 流执行联合控制, 提高了 QoS 管理的灵活性, 满足了不同业务的需求, 提高了用户体验。

应理解, 在上述步骤 S701-S705 中, (1) 第一网元可以为 SMF; (2) 第二网元可以为 UPF; (3) 第一设备可以是终端设备或应用服务器。为了便于理解本申请提供的一种 QoS 管理的方法, 以下, 作为示例而非限定, 以第一网元为 SMF, 第二网元为 UPF, 第一设备为终端设备为例, 参考如图 8 所示的具体示例方法, 分别对方法 700 中的 S701-S705 进行详细说明。

需要说明的是, 下面提到的某些步骤与上述方法 700 相同, 此处对于相关细节不再赘述, 具体过程可参考方法 700 相关步骤。方法 800 以下行数据为例进行说明。

S801, 终端设备向 SMF 发送第三信息, 该第一信息用于为第一业务请求 QoS。

在一种可能的实现方式, 终端设备通过 AMF 向 SMF 发送第一信息。

应理解，终端设备向 SMF 发送第一信息之前，终端设备已经通过 PDU 会话与应用服务器建立连接，则第一信息可以包括在会话修改请求信息中。

5 S802，SMF 发送第四信息，该第四信息用于指示联合控制第二 QoS 流和第三 QoS 流以及用于联合控制该第二 QoS 流和该第三 QoS 流的 QoS 参数。第二 QoS 流和第三 QoS 流用于传输第二业务的数据流。

具体的，SMF 向终端设备、接入网设备、UPF 发送第四信息。

一种可能的实现方式中，SMF 通过 AMF 向接入网设备发送第四信息，通过 AMF 和接入网设备向终端设备发送第四信息。第四信息可以包括在 QoS 配置（QoS profile）中向接入网设备发送，第四信息可以包括在 QoS 规则（QoS rule）中向终端设备发送。第四信
10 息可以包括在包检测规则（Packet detection rule）中向 UPF 设备发送。

S803，应用服务器向 UPF 发送第一业务的数据流。

可选的，在一些实施例中，应用服务器可以在 S803 中指示第二业务的数据流的数据包和数据包组的关系。例如，第二业务的数据流包括 100 个数据包，该 100 个数据包可以组成 3 个数据包组。

15 一种可能的实现方式，应用服务器在第一业务的数据包的包头增加指示信息。

可选的，在一些实施例中，应用服务器还可以在 S803 中指示第一业务的数据流中每一个数据包组的加权因子。加权因子可以对应不同的重要性，即可以理解为对于用户影响的程度。以 XR 业务为例，XR 业务的关键帧对于用户的影响程度大，辅助帧对应用户的影响程度小，因此 XR 业务的关键帧的加权因子大于辅助帧的加权因子。

20 应理解，针对 S803 的描述可以参见 S503 的描述，在此不再赘述。

S804，UPF 向接入网设备发送第二业务的数据流。

具体的，UPF 通过第二 QoS 流和第三 QoS 流传输第二业务的数据流，并根据第四信息对第二 QoS 流和第三 QoS 流执行联合 QoS 控制。

25 可选的，在一些实施例中，UPF 可以在第二业务的数据包的包头中增加标记信息，用于指示第二业务的数据包组的数量和/或每一个数据包组的起始数据包和终止数据包和/或每一个数据包的加权因子和/或每一个数据包组包括的数据包的数量，以使接入网设备和终端设备可以基于数据包组为粒度对第二 QoS 流和第三 QoS 流执行联合 QoS 控制。

一种可能的实现方式，UPF 可以通过应用服务器指示得到该标记信息，即通过 S803 步骤获取。

30 一种可能的实现方式，UPF 可以根据第二业务的业务信息得到该标记信息。

S805，接入网设备向终端设备发送第二业务的数据流。

具体的，接入网设备接收 UPF 发送第二业务的数据，接入网设备根据第二信息对第二业务的数据流执行 QoS 控制。例如，接入网设备可以基于联合错误率调度第二业务的数据包，接入网设备需要保证第二 QoS 流和第三 QoS 流传输错误的数据包（或数据包组）
35 与总数据包（或总数据包组）的比率小于联合错误率。

S806，终端设备根据第二信息对第二业务的数据流执行 QoS 控制。

应理解，方法 800 中以传输下行数据为例，但本申请实施例并不限于此，本申请实施例还可以用于传输上行数据。例如，终端设备可以基于联合错误率调度第二业务的数据包，终端设备需要保证第二 QoS 流和第三 QoS 流传输错误的数据包（或数据包组）与总

数据包（或总数据包组）的比率小于联合错误率。

应理解，当传输上行数据时，终端设备可以类似于应用服务器指示第二业务的数据包与数据包组的关系和/或每一个数据包组的加权因子。

本申请实施的 QoS 管理方法，当业务的数据流承载于多个 QoS 流时，可以对该多个
5 QoS 流执行联合控制，提高了 QoS 管理的灵活性，满足了不同业务的需求，提高了用户体验。

图 9 和图 10 为本申请的实施例提供的可能的 QoS 管理的装置的示意性框图。这些装置可以实现上述方法实施例中终端设备或任一网元的功能，因此也能实现上述方法实施例所具备的有益效果。在本申请实施例中，该装置可以是终端设备，也可以是第一网元，也
10 可以是第二网元、接入网设备，还可以是应用于终端设备、第一网元、第二网元（如芯片）。

图 9 是本申请实施例提供的一种 QoS 管理的装置的示意性框图，该装置 900 包括收发单元 901，可选的，还可以包括处理单元 902。

当装置 900 用于实现图 4 所述方法实施例中第一设备的功能时，该收发单元 901 用于发送第一信息，该第一信息用于为第一业务请求 QoS；该收发单元 901 还用于接收第二信息，该第二信息用于指示第一 QoS 流的控制方式和 QoS 参数，其中第一 QoS 流用于传输
15 第一业务的数据流。处理单元 902 用于根据第二信息对该第一业务的数据流执行 QoS 控制。

当装置 900 用于实现图 4 所述方法实施例中第一网元的功能时，该收发单元 901 用于接收来自第一设备的第一信息，该第一信息用于为第一业务请求 QoS；该收发单元 901 还
20 用于发送第二信息，该第二信息用于指示第一 QoS 流的控制方式和 QoS 参数，其中第一 QoS 流用于传输第一业务的数据流。处理单元 902 用于确定第一 QoS 流的控制方式和 QoS 参数。

当装置 900 用于实现图 4 所述方法实施例中第二网元的功能时，该收发单元 901 用于接收第二信息，该第二信息用于指示第一 QoS 流的控制方式和 QoS 参数，其中第一 QoS
25 流用于传输第一业务的数据流。处理单元 902 用于根据第二信息对该第一业务的数据流执行 QoS 控制。

当装置 900 用于实现图 4 所述方法实施例中接入网设备的功能时，该收发单元 901 用于接收第二信息，该第二信息用于指示第一 QoS 流的控制方式和 QoS 参数，其中第一 QoS
30 流用于传输第一业务的数据流。处理单元 902 用于根据第二信息对该第一业务的数据流执行 QoS 控制。

当装置 900 用于实现图 7 所述方法实施例中第一设备的功能时，该收发单元 901 用于发送第三信息，该第三信息用于为第二业务请求 QoS；该收发单元 901 还用于接收第四信息，该第四信息用于指示联合控制第二 QoS 流和第三 QoS 流以及用于联合控制第二 QoS
35 流和第三 QoS 流的 QoS 参数，其中第二 QoS 流和第三 QoS 流用于传输所述第二业务的数据流。处理单元 902 用于根据第四信息对该第一业务的数据流执行 QoS 控制。

当装置 900 用于实现图 7 所述方法实施例中第一网元的功能时，该收发单元 901 用于接收来自第一设备的第三信息，该第三信息用于为第二业务请求 QoS；该收发单元 901 还用于发送第四信息，该第四信息用于指示联合控制第二 QoS 流和第三 QoS 流以及用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数，其中第二 QoS 流和第三 QoS 流用于传输

所述第二业务的数据流。处理单元 902 用于确定用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数。

当装置 900 用于实现图 7 所述方法实施例中第二网元的功能时，该收发单元 901 用于接收第四信息，该第四信息用于指示联合控制第二 QoS 流和第三 QoS 流以及用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数，其中第二 QoS 流和第三 QoS 流用于传输所述第二业务的数据流。处理单元 902 用于根据第四信息对该第一业务的数据流执行 QoS 控制。

当装置 900 用于实现图 7 所述方法实施例中接入网设备的功能时，该收发单元 901 用于接收第二信息，该第四信息用于指示联合控制第二 QoS 流和第三 QoS 流以及用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数，其中第二 QoS 流和第三 QoS 流用于传输所述第二业务的数据流。处理单元 902 用于根据第四信息对该第一业务的数据流执行 QoS 控制。

关于上述收发单元 901 和处理单元 902 更详细的描述，可参考上述方法实施例 400 至方法 800 中的相关描述，在此不再说明。

图 10 示出了应用本申请实施例的装置 1000 的示意性框图。上述方法 400 至方法 800 中任一方法所涉及的任一网元，都可以由图 10 所示的装置来实现。

应理解，装置 1000 可以是实体设备，也可以是实体设备的部件（例如，集成电路，芯片等等），还可以是实体设备中的功能模块。

如图 10 所示，该装置 1000 包括：一个或多个处理器 1001。处理器 1001 可以存储用于执行本申请实施例的方法的执行指令。可选地，处理器 1001 中可以调用接口实现接收和发送功能。所述接口可以是逻辑接口或物理接口，对此不作限定。例如，接口可以是收发电路，或是接口电路。用于实现接收和发送功能的收发电路、或接口电路可以是分开的，也可以集成在一起。上述收发电路或接口电路可以用于代码/数据的读写，或者，上述收发电路或接口电路可以用于信号的传输或传递。

可选地，接口可以通过收发器实现。可选地，该装置 1000 还可以包括收发器 1003。所述收发器 1003 可以称为收发单元、收发机、收发电路或者收发器等，用于实现收发功能。

可选地，该装置 1000 还可以包括存储器 1002。本申请实施例对存储器 1002 的具体部署位置不作具体限定，该存储器可以集成于处理器中，也可以是独立于处理器之外。对于该装置 1000 不包括存储器的情形，该装置 1000 具备处理功能即可，存储器可以部署在其他位置（如，云系统）。

处理器 1001、存储器 1002 和收发器 1003 之间通过内部连接通路互相通信，传递控制和/或数据信号。

可以理解的是，尽管并未示出，装置 1000 还可以包括其他装置，例如输入装置、输出装置、电池等。

可选的，在一些实施例中，存储器 1002 可以存储用于执行本申请实施例的方法的执行指令。处理器 1001 可以执行存储器 1002 中存储的指令结合其他硬件（例如收发器 703）完成下文所示方法执行的步骤，具体工作过程和有益效果可以参见下文方法实施例中的描述。

本申请实施例揭示的方法可以应用于处理器 1001 中，或者由处理器 1001 实现。处理器 1001 可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，方法的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器可以是通用处理器、数字信号处理器（digital signal processor, DSP）、专用集成电路（application specific integrated circuit, ASIC）、现场可编程门阵列（field programmable gate array, FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存取存储器（random access memory, RAM）、闪存、只读存储器（read-only memory, ROM）、可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器，处理器读取存储器中的指令，结合其硬件完成上述方法的步骤。

可以理解，存储器 1002 可以是易失性存储器或非易失性存储器，或可包括易失性和非易失性存储器两者。其中，非易失性存储器可以是只读存储器 ROM、可编程只读存储器（programmable ROM, PROM）、可擦除可编程只读存储器（erasable PROM, EPROM）、电可擦除可编程只读存储器（electrically EPROM, EEPROM）或闪存。易失性存储器可以是随机存取存储器 RAM，其用作外部高速缓存。通过示例性但不是限制性说明，许多形式的 RAM 可用，例如静态随机存取存储器（static RAM, SRAM）、动态随机存取存储器（dynamic RAM, DRAM）、同步动态随机存取存储器（synchronous DRAM, SDRAM）、双倍数据速率同步动态随机存取存储器（double data rate SDRAM, DDR SDRAM）、增强型同步动态随机存取存储器（enhanced SDRAM, ESDRAM）、同步连接动态随机存取存储器（synchlink DRAM, SLDRAM）和直接内存总线随机存取存储器（direct rambus RAM, DR RAM）。应注意，本文描述的系统和方法的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

另外，在本申请中，装置 900 是以功能模块的形式来呈现。这里的“模块”可以指特定应用集成电路 ASIC、电路、执行一个或多个软件或固件程序的处理器和存储器、集成逻辑电路，和/或其他可以提供上述功能的器件。在一个简单的实施例中，本领域的技术人员可以想到装置 900 可以采用图 9 所示的形式。处理单元 901 可以通过图 10 所示的处理器 1001 来实现。可选地，如果图 10 所示的计算机设备包括存储器 1002，处理单元 902 可以通过处理器 1001 和存储器 1002 来实现。收发单元 901 可以通过图 10 所示的收发器 1003 来实现。所述收发器 1003 包括接收功能和发送功能。具体的，处理器通过执行存储器中存储的计算机程序来实现。可选地，当所述装置 900 是芯片时，那么收发单元 1001 的功能和/或实现过程还可以通过管脚或电路等来实现。可选地，所述存储器可以为所述芯片内的存储单元，比如寄存器、缓存等，所述存储单元还可以是所述计算机设备内的位于所述芯片外部的存储单元，如图 10 所示的存储器 1002，或者，也可以是部署在其他系统或设备中的存储单元，不在所述计算机设备内。本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的

特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

本申请的各个方面或特征可以实现成方法、装置或使用标准编程和/或工程技术的制品。本申请中使用的术语“制品”涵盖可从任何计算机可读器件、载体或介质访问的计算机程序。例如，计算机可读介质可以包括，但不限于：磁存储器件（例如，硬盘、软盘或磁带等），光盘（例如，压缩盘（compact disc，CD）、数字通用盘（digital versatile disc，DVD）等），智能卡和闪存器件（例如，可擦写可编程只读存储器（erasable programmable read-only memory，EPROM）、卡、棒或钥匙驱动器等）。另外，本文描述的各种存储介质可代表用于存储信息的一个或多个设备和/或其它机器可读介质。术语“机器可读介质”可包括但不限于，无线信道和能够存储、包含和/或承载指令和/或数据的各种其它介质。

本申请还提供了一种计算机可读介质，其上存储有计算机程序，该计算机程序被计算机执行时实现上述任一方法实施例的功能。

本申请还提供了一种计算机程序产品，该计算机程序产品被计算机执行时实现上述任一方法实施例的功能。在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机指令时，全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线（digital subscriber line，DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质（例如，软盘、硬盘、磁带）、光介质（例如，高密度数字视频光盘（digital video disc，DVD））、或者半导体介质（例如，固态硬盘（solid state disk，SSD））等。

应理解，说明书通篇中提到的“实施例”意味着与实施例有关的特定特征、结构或特性包括在本申请的至少一个实施例中。因此，在整个说明书各个实施例未必一定指相同的实施例。此外，这些特定的特征、结构或特性可以任意适合的方式结合在一个或多个实施例中。应理解，在本申请的各种实施例中，上述各过程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。

还应理解，在本申请中，“当...时”、“若”以及“如果”均指在某种客观情况下UE或者基站会做出相应的处理，并非是限定时间，且也不要求UE或基站实现时一定要有的判断的动作，也不意味着存在其它限定。

另外，本文中术语“系统”和“网络”在本文中常被可互换使用。本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。

本文中术语“...中的至少一个”或“...中的至少一种”，表示所列出的各项的全

部或任意组合，例如，“A、B和C中的至少一种”，可以表示：单独存在A，单独存在B，单独存在C，同时存在A和B，同时存在B和C，同时存在A、B和C这六种情况。

本申请中，若无特殊说明，“至少一个”指一个或多个，“多个”指两个或大于两个。

应理解，在本申请各实施例中，“与A相应的B”表示B与A相关联，根据A可以确定B。但还应理解，根据A确定B并不意味着仅仅根据A确定B，还可以根据A和/或其它信息确定B。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，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单元中。

所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本申请的具体实施方式，但本申请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申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

权利要求书

1. 一种 QoS 管理方法, 其特征在于, 所述方法包括:
第一设备发送第一信息, 所述第一信息用于为第一业务请求 QoS;
5 所述第一设备接收第二信息, 所述第二信息用于指示第一 QoS 流的 QoS 控制方式和 QoS 参数, 其中所述第一 QoS 流用于传输所述第一业务的数据流, 所述 QoS 控制方式包括基于数据包组控制的第一控制方式;
所述第一设备根据所述第二信息对所述第一业务的数据流执行 QoS 控制。
2. 根据权利要求 1 所述的方法, 其特征在于, 所述方法还包括:
10 所述 QoS 控制方式还包括基于数据包控制的第二控制方式。
3. 根据权利要求 1 或 2 所述的方法, 其特征在于, 所述第二信息包括第一 QoS 参数, 所述第一 QoS 参数包括 N 个基于数据包组的 QoS 参数, $N \geq 1$ 且 N 为正整数。
4. 根据权利要求 1 所述的方法, 其特征在于, 所述第二信息包括第一指示信息和第二 QoS 参数, 所述第一指示信息用于指示所述第一 QoS 流的 QoS 控制方式, 所述第二
15 QoS 参数用于指示所述第一 QoS 流的 QoS 参数。
5. 根据权利要求 4 所述的方法, 其特征在于, 当所述第一指示信息指示所述第一 QoS 流的 QoS 控制方式为第一控制方式时, 所述第二 QoS 参数包括至少一个基于数据包组的 QoS 参数; 或,
当所述第二指示信息指示所述第一 QoS 流的 QoS 控制方式为第二控制方式时, 所述
20 第二 QoS 参数包括至少一个基于数据包的 QoS 参数。
6. 根据权利要求 4 所述的方法, 其特征在于, 所述第二 QoS 参数包括至少一个基于数据包的 QoS 参数。
7. 根据权利要求 4 至 6 中任一项所述的方法, 其特征在于, 所述第一指示信息用于指示所述第一 QoS 流的优选 QoS 控制方式。
- 25 8. 根据权利要求 1 至 7 中任一项所述的方法, 其特征在于, 所述方法还包括: 所述第一设备接收所述第一业务的业务信息, 所述业务信息用于确定所述第一 QoS 流对应的 QoS 参数。
9. 根据权利要求 8 所述的方法, 其特征在于, 所述业务信息包括所述第一业务的编码参数和/或所述第一业务中每一个数据包组的加权因子。
- 30 10. 根据权利要求 9 所述的方法, 其特征在于, 所述方法还包括:
所述第一设备根据所述第二信息和所述业务信息确定所述第一 QoS 流的 QoS 参数。
11. 根据权利要求 1 所述的方法, 其特征在于, 所述第二信息包括第二指示信息、第三 QoS 参数和第四 QoS 参数, 所述第二指示信息用于指示所述第一 QoS 流的优选 QoS 控制方式, 所述第三 QoS 参数包括 M 个基于数据包组的 QoS 参数, $M \geq 1$ 且 M 为正整数,
35 所述第四 QoS 参数包括 L 个基于数据包的 QoS 参数, $L \geq 1$ 且 L 为正整数。
12. 根据权利要求 3 所述的方法, 其特征在于, 所述第一 QoS 参数包括组优先级、组延迟预算、组错误率、聚合组错误率、最大组丢失率和最大聚合组丢失率中的至少一种。
13. 根据权利要求 12 所述的方法, 其特征在于, 所述组优先级用于指示所述第一 QoS

流中不同数据包组调度的优先级;

所述组延迟预算用于指示所述第一 QoS 流中数据包组在所述第一设备与第二网元之间的时延上限, 其中第二网元为用户面网元 UPF;

所述组错误率用于指示所述第一 QoS 流中未成功传送的数据包组的上限;

5 所述聚合组错误率用于指示所述第一 QoS 流中未成功传送的加权数据包组的上限;

所述最大组丢失率用于指示所述第一 QoS 流中可容忍的丢弃的数据包组的上限;

所述最大聚合组丢失率用于指示所述第一 QoS 流中可容忍的丢弃的加权数据包组的上限。

14. 一种 QoS 管理方法, 其特征在于, 所述方法包括:

10 第一网元接收第一设备发送的第一信息, 所述第一信息用于为第一业务请求 QoS;

所述第一网元发送第二信息, 所述第二信息用于指示第一 QoS 流的 QoS 控制方式和 QoS 参数, 其中所述第一 QoS 流用于传输所述第一业务的数据流, 所述 QoS 控制方式包括基于数据包组控制的第一控制方式。

15 15. 根据权利要求 14 所述的方法, 其特征在于, 所述方法还包括:

15 所述 QoS 控制方式还包括基于数据包控制的第二控制方式。

16. 根据权利要求 14 或 15 所述的方法, 其特征在于, 所述第二信息包括第一 QoS 参数, 所述第一 QoS 参数包括 N 个基于数据包组的 QoS 参数, $N \geq 1$ 且 N 为正整数。

17. 根据权利要求 14 所述的方法, 其特征在于, 所述第二信息包括第一指示信息和第二 QoS 参数, 所述第一指示信息用于指示所述第一 QoS 流的 QoS 控制方式, 所述第二 QoS 参数用于指示所述第一 QoS 流的 QoS 参数。

18. 根据权利要求 17 所述的方法, 其特征在于, 当所述第一指示信息指示所述第一 QoS 流的 QoS 控制方式为第一控制方式时, 所述第二 QoS 参数包括至少一个基于数据包组的 QoS 参数; 或,

25 当所述第二指示信息指示所述第一 QoS 流的 QoS 控制方式为第二控制方式时, 所述第二 QoS 参数包括至少一个基于数据包组的 QoS 参数。

19. 根据权利要求 17 所述的方法, 其特征在于, 所述第二 QoS 参数包括至少一个基于数据包组的 QoS 参数。

20. 根据权利要求 17 至 19 中任一项所述的方法, 其特征在于, 所述第一指示信息具体用于指示所述第一 QoS 流的优选 QoS 控制方式。

30 21. 根据权利要求 14 至 20 中任一项所述的方法, 其特征在于, 所述方法还包括: 所述第一网元获取所述第一业务的业务信息, 所述业务信息用于确定所述第一 QoS 流对应的 QoS 参数。

22. 根据权利要求 21 所述的方法, 其特征在于, 所述业务信息包括所述第一业务的编码参数和/或所述第一业务中每一个数据包组的加权因子。

35 23. 根据权利要求 21 或 22 所述的方法, 其特征在于, 所述方法还包括:

所述第一网元发送所述业务信息。

24. 根据权利要求 14 所述的方法, 其特征在于, 所述第二信息包括第二指示信息、第三 QoS 参数和第四 QoS 参数, 所述第二指示信息用于指示所述第一 QoS 流的优选 QoS 控制方式, 所述第三 QoS 参数包括 M 个基于数据包组的 QoS 参数, $M \geq 1$ 且 M 为正整数,

所述第四 QoS 参数包括 L 个基于数据包的 QoS 参数, $L \geq 1$ 且 L 为正整数。

25. 根据权利要求 14 至 24 中任一项所述的方法, 其特征在于, 所述第一网元发送所述第二信息之前, 所述方法还包括:

所述第一网元根据所述第一业务确定所述第一 QoS 流的 QoS 控制方式。

5 26. 根据权利要求 14 至 24 中任一项所述的方法, 其特征在于, 所述第一网元发送所述第二信息之前, 所述方法还包括:

所述第一网元根据所述第一设备是否支持第一控制方式确定所述第一 QoS 流的 QoS 控制方式。

27. 根据权利要求 16 所述的方法, 其特征在于, 所述第一 QoS 参数包括组优先级、
10 组延迟预算、组错误率、聚合组错误率、最大组丢失率和最大聚合组丢失率中的至少一种。

28. 根据权利要求 27 所述的方法, 其特征在于, 所述组优先级用于指示所述第一 QoS 流中不同数据包组调度的优先级;

所述组延迟预算用于指示所述第一 QoS 流中数据包组在所述第一设备与第二网元之间的时延上限, 其中第二网元为用户面网元 UPF;

15 所述组错误率用于指示所述第一 QoS 流中未成功传送的数据包组的上限;

所述聚合组错误率用于指示所述第一 QoS 流中未成功传送的加权数据包组的上限;

所述最大组丢失率用于指示所述第一 QoS 流中可容忍的丢弃的数据包组的上限;

所述最大聚合组丢失率用于指示所述第一 QoS 流中可容忍的丢弃的加权数据包组的上限。

20 29. 一种 QoS 管理方法, 其特征在于, 所述方法包括:

第一设备发送第三信息, 所述第三信息用于为第二业务请求 QoS;

所述第一设备接收第四信息, 所述第四信息包括用于联合控制第二 QoS 流和第三 QoS 流的 QoS 参数, 其中所述第二 QoS 流和第三 QoS 流用于传输所述第二业务的数据流;

所述第一设备根据所述第四信息对所述第二业务的数据流执行联合 QoS 控制。

25 30. 根据权利要求 29 所述的方法, 其特征在于, 所述第四信息包括第五 QoS 参数, 所述第五 QoS 参数包括 K 个联合 QoS 参数, $K \geq 1$ 且 K 为正整数。

31. 根据权利要求 29 所述的方法, 其特征在于, 所述第四信息包括第三指示信息和第六 QoS 参数, 所述第三指示信息用于指示联合控制所述第二 QoS 流和所述第三 QoS 流, 所述第六 QoS 参数为用于联合 QoS 控制的参数, 所述第六 QoS 包括 L 个基于数据包组的
30 QoS 参数, $L \geq 1$ 且 L 为正整数; 或所述第六 QoS 参数包括 J 个基于数据包的 QoS 参数, $J \geq 1$ 且 J 为正整数。

32. 根据权利要求 30 所述的方法, 其特征在于, 所述第五 QoS 参数包括: 联合错误率、聚合联合错误率、联合最大丢失率、聚合联合最大丢失率、联合最大比特速率中的至少一种。

35 33. 根据权利要求 32 所述的方法, 其特征在于, 所述联合错误率用于指示所述第二 QoS 流和所述第三 QoS 流中未成功传送的数据包组或数据包的上限;

所述聚合联合错误率用于指示所述第二 QoS 流和所述第三 QoS 流中未成功传送的加权数据包组或数据包的上限;

所述联合最大丢失率用于指示所述第二 QoS 流和所述第三 QoS 流中可容忍的丢弃的

数据包组或数据包的上限;

所述聚合联合最大组丢失率用于指示所述第二 QoS 流和所述第三 QoS 流中可容忍的丢弃的加权数据包组或加权数据包的上限;

5 所述联合最大比特速率用于指示所述第二 QoS 流和所述第三 QoS 流的最大聚合比特速率的上限。

34. 一种 QoS 管理方法, 其特征在于, 所述方法包括:

第一网元接收第一设备发送的第三信息, 所述第三信息用于为第二业务请求 QoS;

所述第一网元发送第四信息, 所述第四信息包括联合控制第二 QoS 流和第三 QoS 流的 QoS 参数, 其中所述第二 QoS 流和第三 QoS 流用于传输所述第二业务的数据流。

10 35. 根据权利要求 34 所述的方法, 其特征在于, 所述第四信息包括第五 QoS 参数, 所述第五 QoS 参数包括 K 个联合 QoS 参数, $K \geq 1$ 且 K 为正整数。

36. 根据权利要求 34 所述的方法, 其特征在于, 所述第四信息包括第三指示信息和第六 QoS 参数, 所述第三指示信息用于指示联合控制所述第二 QoS 流和所述第三 QoS 流, 所述第六 QoS 参数为用于联合 QoS 控制的参数, 所述第六 QoS 包括 L 个基于数据包组的 QoS 参数, $L \geq 1$ 且 L 为正整数; 或所述第六 QoS 参数包括 J 个基于数据包的 QoS 参数, $J \geq 1$ 且 J 为正整数。

37. 根据权利要求 35 所述的方法, 其特征在于, 所述第五 QoS 参数包括: 联合错误率、聚合联合错误率、联合最大丢失率、聚合联合最大丢失率、联合最大比特速率中的至少一种。

20 38. 根据权利要求 37 所述的方法, 其特征在于, 所述联合错误率用于指示所述第二 QoS 流和所述第三 QoS 流中未成功传送的数据包组或数据包的上限;

所述聚合联合错误率用于指示所述第二 QoS 流和所述第三 QoS 流中未成功传送的加权数据包组或数据包的上限;

25 所述联合最大丢失率用于指示所述第二 QoS 流和所述第三 QoS 流中可容忍的丢弃的数据包组或数据包的上限;

所述聚合联合最大组丢失率用于指示所述第二 QoS 流和所述第三 QoS 流中可容忍的丢弃的加权数据包组或加权数据包的上限;

所述联合最大比特速率用于指示所述第二 QoS 流和所述第三 QoS 流的最大聚合比特速率的上限。

30 39. 一种装置, 其特征在于, 包括用于执行权利要求 1 至 38 中任一项所述的方法的模块或单元。

40. 一种装置, 其特征在于, 包括处理器, 所述处理器, 用于执行存储器中存储的计算机程序或指令, 以使得所述装置执行权利要求 1 至 38 中任一项所述的方法。

35 41. 一种计算机可读存储介质, 其特征在于, 所述计算机可读存储介质上存储有计算机程序或指令, 当所述计算机程序或指令在计算机上运行时, 使得所述计算机执行如权利要求 1 至 38 中任一项所述的方法。

42. 一种计算机程序产品, 其特征在于, 所述计算机程序产品包括用于执行如权利要求 1 至 38 中任一项所述的方法的计算机程序或指令。

43. 一种芯片, 其特征在于, 所述芯片与存储器耦合, 用于读取并执行所述存储器中

存储的程序指令，以实现如权利要求 1 至 38 中任一项所述的方法。