[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04L 12/56 (2006.01)
H04L 1/00 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200810216948.5

[43] 公开日 2009 年 6 月 24 日

[11] 公开号 CN 101465806A

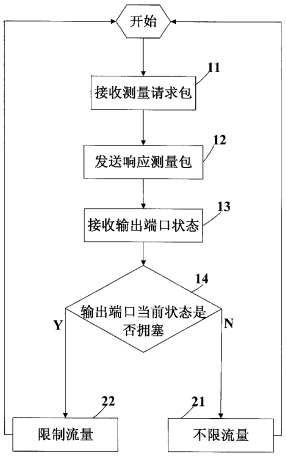
[22] 申请日 2008.10.22
[21] 申请号 200810216948.5
[71] 申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
[72] 发明人 陈武茂 伊学文
权利要求书 5 页 说明书 20 页 附图 10 页

[54] 发明名称

一种调度交换网数据包的方法、装置和系统

[57] 摘要

本发明涉及网络通信领域，尤其涉及一种管理交换网的方法和系统。 本发明提供一种调度交换网数据包的方法、装置和系统。 该方法主要包括：接收交换网输出端口判断的输出端口状态，所述输出端口状态是对包的转发延时测量结果的判断；根据所述输出端口状态管理所述交换网输出端口对应虚拟输出队列 VoQ 的最大流量。 该装置主要包括：VoQ 状态维护模块；测量包发送模块；VoQ 流量限制模块。 一种交换网输出端口检控装置，包括：输出端口状态维护发布模块；状态判断模块；请求包发送模块。 该系统主要包括：交换网输出端口检控装置；交换网输入端口管理装置。



1、一种调度交换网数据包的方法，其特征在于，包括:

接收交换网输出端口判断的输出端口状态，所述输出端口状态是对包的转发延时测量结果的判断;

根据所述输出端口状态管理所述交换网输出端口对应虚拟输出队列VoQ的最大流量。

2、根据权利要求1所述的方法，其特征在于，所述输出端口状态是对包的转发延时测量结果的判断，具体包括:

当所述输出端口的当前状态处于正常状态且所述包的转发延时测量结果大于预设的拥塞发生判断域值时，输出端口状态改变为拥塞状态的判断;

当所述输出端口的当前状态处于拥塞状态且所述包的转发延时测量结果小于预设的拥塞解除判断域值时，输出端口状态改变为正常状态的判断。

3、根据权利要求2所述的方法，其特征在于，所述接收交换网输出端口判断的输出端口状态，所述输出端口状态是对包的转发延时测量结果的判断，具体包括:

接收所述交换网输出端口由所述交换网的最高优先级发送的测量请求包;

收到所述测量请求包后由所述交换网的普通优先级向所述交换网输出端口发送测量包;

接收所述交换网输出端口判断的输出端口状态;

所述交换网输出端口是根据收到所述测量包的时刻减去所述测量请求包的发送时刻的差值作为所述包的转发延时判断输出端口状态。

4、根据权利要求3所述的方法，其特征在于，所述接收交换网输出端口判断的输出端口状态，所述输出端口状态是对包的转发延时测量结果的判断，还可以包括:

由所述交换网的最高优先级向所述交换网输出端口发送最高优先级测量包;

由所述交换网的普通优先级向所述交换网输出端口发送普通测量包;

接收所述交换网输出端口判断的输出端口状态;

所述交换网输出端口是根据收到所述普通测量包的时刻减去收到所述最高优先级测量包的时刻的差值作为所述包的转发延时判断输出端口状态。

5、根据权利要求3所述的方法，其特征在于，当所述交换网支持全局时钟同步时，所述接收交换网输出端口判断的输出端口状态，所述输出端口状态是对包的转发延时测量结果的判断，还可以包括:

向所述交换网输出端口发送打上进入交换网时戳的测量包;

接收所述交换网输出端口判断的输出端口状态;

所述交换网输出端口是根据收到所述打上交换网时戳的测量包的本地时刻减去所述测量包的进入交换网时戳作为所述包的转发延时判断输出端口状态。

6、根据权利要求5所述的方法，其特征在于，根据所述输出端口的状态管理所述交换网输入端口对应VoQ的最大流量，具体包括:

当所述输出端口的状态处于所述正常状态时，对所述交换网输出端口对应VoQ不做流量限制;

当所述输出端口的状态处于所述拥塞状态时，对所述交换网输出端口对应VoQ的最大流量限制为交换网接口带宽除以交换网入口的总数。

7、一种交换网数据包调度装置，其特征在于，包括:

VoQ状态维护模块，用于维护VoQ状态信息，并接收根据对包的转发延时进行测量的结果判断的输出端口状态，根据所述输出端口状态更新所述VoQ状态信息;

测量包发送模块，用于根据交换网输出端口的请求，发送测量包;

VoQ流量限制模块，用于根据所述VoQ状态维护模块的所述VoQ状态信息管理VoQ流量。

8、根据权利要求7所述的装置，其特征在于，所述测量包发送模块还用于:

当所述VoQ状态维护模块的所述VoQ状态信息是正常状态时，发送所述测量包。

9、一种交换网输出端口检控装置，其特征在于，包括:

输出端口状态维护发布模块，用于维护输出端口状态信息，并向交换网输入端口发布所述输出端口状态信息，所述输出端口状态信息包括正常状态和拥塞状态;

状态判断模块，用于计算包的转发延时，并根据所述包的转发延时判断输出端口状态信息，并更新所述输出端口状态维护发布模块维护的所述输出端口状态信息;

请求包发送模块，用于发送测量请求包，并向状态判断模块通告所述测量请求包的发送时刻。

10、一种调度交换网数据包的系统，其特征在于，包括:

交换网输出端口检控装置，用于根据测量包的延时检测发布输出端口的状态;

交换网输入端口管理装置，用于发送所述测量包并根据所述交换网输出端口检控装置发布的所述输出端口的状态管理VoQ。

11、根据权利要求10所述的系统，其特征在于，所述交换网输入端口管理装置具体包括:

VoQ状态维护模块，用于维护VoQ状态信息，并根据所述交换网输入端口管理装置发布的所述输出端口的状态更新所述VoQ状态信息，所述VoQ状态信息包括正常状态和拥塞状态;

测量包发送模块，用于根据所述交换网输出端口管理装置的请求，发送所述测量包;

VoQ流量限制模块，用于根据所述VoQ状态维护模块的所述VoQ状态信息管理VoQ流量。

12、根据权利要求11所述的系统，其特征在于，所述测量包发送模块还用于:

当所述VoQ状态维护模块的所述VoQ状态信息是正常状态时，发送所述测量包。

13、根据权利要求10所述的系统，其特征在于，所述交换网输出端口检控装置具体包括:

输出端口状态维护发布模块，用于维护输出端口状态信息，并向所述交换网输入端口管理装置发布所述输出端口状态信息，所述输出端口状态信息包括正常状态和拥塞状态;

状态判断模块，用于计算包的转发延时，并根据所述包的转发延时判断输出端口状态信息，并更新所述输出端口状态维护发布模块维护的所述输出端口状态信息;

请求包发送模块，用于发送测量请求包，并向状态判断模块通告所述测量请求包的发送时刻。

一种调度交换网数据包的方法、装置和系统

技术领域

本发明涉及网络通信领域，尤其涉及一种管理交换网的方法、装置和系统。

背景技术

目前大容量、高吞吐率的分组交换网通常都会采用输出缓存结构（在交换网的输出端口缓存流量）。

这是因为，在很多情况下，不带输出缓存的交换网的吞吐率是比较低的。例如，对一个不带输出缓存的交换网，当两个输入端口同时有流量指向同一个输出端口时，由于交换网只能允许其中一个输入端口发送流量，另一个输入端口只能等待，这时将有一个输出端口处于空闲状态。

理论上，具有无限大缓存空间的输出缓存交换网可以保证在任何流量下，任何一个输入端口在任何时候都可以发送流量，只要输入端口有流量指向某个输出端口，该输出端口就不可能为空闲。

但是，交换网内部的缓存空间不可能做到无穷大。在这种情况下，如果输入端口采用普通的先进先出FIFO（first in first out）队列时，若遇到突发度很高的流量，不对输入进行反压控制会导致分组丢失，而进行反压控制又会影响交换网的吞吐率，这是因为FIFO队列输入会造成队头阻塞（head-of-line blocking）。例如，由于流量突发，交换网的1号输出端口发生拥塞，为了避免包丢失，交换网通知所有输入端口不要再向交换网发送目的端口为1号的分组。这时，假如某个输入端口的FIFO中第一个分组的目的端口是1号，则它将一直处于等待状态，直到1号目的端口解除拥塞。而FIFO中的其他分组尽管目的端口不是1号，也无法得到处理。也就是说，这段时间内该输入端

口虽然有一些流量需要发往交换网其他的没有拥塞的端口，但由于队头阻塞，这些流量在这段时间不能通过交换网，这样交换网的吞吐率就下降了。

为了克服队头阻塞问题，可以在每一个输入端口使用N（输出端口数）个分立的队列，每个队列对应一个输出端口，这就是虚拟输出队列VoQ（Virtual Output Queuing）。

交换网检测输出缓存的使用情况，将各个输出队列的反压信息（哪些输出队列已满、哪些仍然可以接收流量）传递给VoQ调度器，这样VoQ调度器就可以决定应该选择哪个VoQ发送，从而避免了队头阻塞。

显然，只有当输入端口的VoQ调度器知道哪个输出队列已满、哪个仍可以接收信元时，VOQ 才可以有效地工作。假如交换网不支持传递反压信息（例如使用LSW（Lan switch，局域网交换）做交换网）；或者虽然支持，但由于交换网和VoQ调度器是由不同厂家设计的，消息格式不兼容；则VoQ调度器无法正常工作。

发明内容

本发明实施例的目的是提供一种交换网输入端口的VoQ调度器无法接收信元的情况下，调度交换网数据包的方法、装置和系统。

本发明实施例的目的是通过以下技术方案实现的：

一种调度交换网数据包的方法，包括：

接收交换网输出端口判断的输出端口状态，所述输出端口状态是对包的转发延时测量结果的判断；

根据所述输出端口状态管理所述交换网输出端口对应虚拟输出队列VoQ的最大流量。

一种交换网数据包调度装置，包括：

VoQ状态维护模块，用于维护VoQ状态信息，并接收根据对包的转发延时进行测量的结果判断的输出端口状态，根据所述输出端口状态更新所述VoQ状态信息；

测量包发送模块，用于根据交换网输出端口的请求，发送测量包；

VoQ流量限制模块，用于根据所述VoQ状态维护模块的所述VoQ状态信息管理VoQ流量。

一种交换网输出端口检控装置，包括：

输出端口状态维护发布模块，用于维护输出端口状态信息，并向交换网输入端口发布所述输出端口状态信息，所述输出端口状态信息包括正常状态和拥塞状态；

状态判断模块，用于计算包的转发延时，并根据所述包的转发延时判断输出端口状态信息，并更新所述输出端口状态维护发布模块维护的所述输出端口状态信息；

请求包发送模块，用于发送测量请求包，并向状态判断模块通告所述测量请求包的发送时刻。

一种调度交换网数据包的系统，包括：

交换网输出端口检控装置，用于根据测量包的延时检测发布输出端口的状态；

交换网输入端口管理装置，用于发送所述测量包并根据所述交换网输出端口检控装置发布的所述输出端口的状态管理VoQ。

采用本发明实施例提供的技术方案因为无需使用反压信息通告VoQ调度器，能在交换网不支持传递反压信息或者交换网和VoQ调度器消息格式不兼容的情况下解决队头拥塞的问题。

附图说明

图1为本发明实施例所述调度交换网数据包的方法流程图;

图2为本发明实施例所述一种交换网数据包调度装置的框图;

图3为本发明实施例所述一种交换网输出端口检控装置的框图;

图4为本发明实施例所述一种调度交换网数据包的系统框图;

图5为本发明另一实施例所述调度交换网数据包的方法的流程图;

图6为本发明另一实施例所述一种交换网数据包调度装置的框图;

图7为本发明另一实施例所述一种交换网输出端口检控装置的框图;

图8为本发明另一实施例所述一种调度交换网数据包的系统框图;

图9为本发明另一实施例所述在交换网支持全局时钟同步时，调度交换网数据包的方法的流程图;

图10为本发明另一实施例所述在交换网支持全局时钟同步时，一种交换网数据包调度装置的框图;

图11为本发明另一实施例所述在交换网支持全局时钟同步时，一种交换网输出端口检控装置的框图;

图12为本发明另一实施例所述在交换网支持全局时钟同步时，一种调度交换网数据包的系统框图;

图13为本发明实施例场景的示意图。

具体实施方式

以下结合图1、图2、图3和图4提供本发明的一个实施例:

如图1所示，为本发明实施例的调度交换网数据包的方法流程图。该实施例提供了调度交换网数据包的方法，包括:

11、接收测量请求包，接收交换网输出端口由所述交换网的最高优先级

发送的测量请求包;

12、发送响应测量包，收到测量请求包后由交换网的普通优先级发送响应测量包;

13、接收输出端口状态，接收所述交换网输出端口判断的输出端口状态;

在本发明实施例具体场景中，所述交换网输出端口判断输出端口状态是根据收到所述测量包的时刻减去所述测量请求包的发送时刻的差值作为所述包的转发延时判断的:

当所述输出端口的当前状态处于正常状态且所述包的转发延时大于预设的拥塞发生判断域值时，所述交换网输出端口判断它的状态改变为拥塞状态;

当所述输出端口的当前状态处于拥塞状态且所述包的转发延时小于预设的拥塞解除判断域值时，所述交换网输出端口判断它的状态改变为正常状态;

14、输出端口当前状态是否拥塞：根据13接收的输出端口状态判断输出端口当前状态是否拥塞，如果拥塞则执行22，否则执行21;

21、不限流量，当所述输出端口的当前状态处于正常状态时，对所述交换网输出端口对应VoQ不做流量限制;

22、限制流量，当所述输出端口的当前状态处于拥塞状态时，对所述交换网输出端口对应VoQ的最大流量限制为交换网接口带宽除以交换网入口的总数。

如图2所示，为本发明实施例的交换网数据包调度装置框图。该实施例提供了调度交换网数据包的装置。该装置包括：VoQ状态维护模块101、测量包发送模块102和VoQ流量限制模块103，其中:

VoQ状态维护模块101，用于维护每个VoQ的状态信息。VoQ的状态有两种：正常状态和拥塞状态。VoQ状态维护模块101接收根据对包的转发延时进行测量的结果判断的输出端口状态，根据所述输出端口状态更新所述VoQ状态信息。当某个VoQ处于正常状态时，假如接收到对应输出端口的拥塞状态消息，则进入拥塞状态；否则保持正常状态。当某个VoQ处于拥塞状态时，假如接收到对应输出端口的正常状态消息，则进入正常状态；否则保持拥塞状态。

测量包发送模块102，用于当接收到关于某个输出端口的请求测量包后，从普通优先级发一个到该端口的测量包。

VoQ流量限制模块103，用于根据VoQ状态维护模块101的信息管理VoQ流量，对于所有处于正常状态的VoQ，不做流量限制。对于所有处于拥塞状态的VoQ，限制其最大流量为交换网接口带宽除以交换网入口的总数。

如图3所示，为本发明实施例的交换网输出端口检控装置框图。该装置包括：输出端口状态维护发布模块201、请求包发送模块202和状态判断模块203，其中：

输出端口状态维护发布模块201，用于维护输出端口状态信息，输出端口的状态有两种：正常状态和拥塞状态。在输出端口的状态发生变化时（从正常状态变成拥塞状态，或者从拥塞状态变成正常状态），输出端口状态维护发布模块201向各个输入端口发布新的输出端口状态信息；或者输出端口状态维护发布模块201定期将当前输出端口状态信息发布给所有输入端口；或者输出端口状态维护发布模块201既在输出端口的状态发生变化时向各个输入端口发布新的输出端口状态信息，也定期将当前输出端口状态信息发布给所有输入端口。

请求包发送模块202，用于定时发送测量请求包，并向状态判断模块203通告所述测量请求包的发送时刻。

状态判断模块203，用于根据包的转发延时判断输出端口状态信息，以收到交换网输入端口发送的测量包的时刻减去请求包发送模块202通告的时刻作为包的转发延时。当输出端口状态信息处于正常状态且包的转发延时大于预设的拥塞发生判断域值时，状态判断模块203判断输出端口状态信息是拥塞状态；当输出端口状态信息处于拥塞状态且包的转发延时小于预设的拥塞解除判断域值时，状态判断模块203判断输出端口状态信息是正常状态。状态判断模块203判断成功后，使输出端口状态维护发布模块201更新输出端口状态信息。

如图4所示，为本发明实施例的调度交换网数据包的系统框图。该实施例提供了调度交换网数据包的系统。该系统包括:

交换网输入端口管理装置1，用于发送测量包并管理VoQ;

交换网输出端口检控装置2，用于根据测量包的延时检测并发布输出端口的状态。

交换网输入端口管理装置1包括：VoQ状态维护模块101、测量包发送模块102和VoQ流量限制模块103，其中:

VoQ状态维护模块101，用于维护每个VoQ的状态信息。VoQ的状态有两种：正常状态和拥塞状态。VoQ状态维护模块101根据交换网输出端口检控装置2发布的输出端口状态信息维护VoQ的状态信息。当某个VoQ处于正常状态时，假如接收到对应输出端口的拥塞状态消息，则进入拥塞状态；否则保持正常状态。当某个VoQ处于拥塞状态时，假如接收到对应输出端口的正常状态消息，则进入正常状态；否则保持拥塞状态。

测量包发送模块102，用于当接收到交换网输出端口检控装置2的关于某个输出端口的请求测量包后，从普通优先级发一个到该端口的测量包。

VoQ流量限制模块103，用于根据VoQ状态维护模块101的信息管理VoQ

流量，对于所有处于正常状态的VoQ，不做流量限制。对于所有处于拥塞状态的VoQ，限制其最大流量为交换网接口带宽除以交换网入口的总数。

交换网输出端口检控装置2包括：输出端口状态维护发布模块201、请求包发送模块202和状态判断模块203：其中：

输出端口状态维护发布模块201，用于维护输出端口状态信息，输出端口的状态有两种：正常状态和拥塞状态。在输出端口的状态发生变化时（从正常状态变成拥塞状态，或者从拥塞状态变成正常状态），输出端口状态维护发布模块201向各个VoQ状态维护模块101发布新的输出端口状态信息；或者输出端口状态维护发布模块201定期将当前输出端口状态信息发布给所有入口的VoQ状态维护模块101；或者输出端口状态维护发布模块201既在输出端口的状态发生变化时向各个VoQ状态维护模块101发布新的输出端口状态信息，也定期将当前输出端口状态信息发布给所有入口的VoQ状态维护模块101。

请求包发送模块202，用于定时发送测量请求包，并向状态判断模块203通告所述测量请求包的发送时刻。

状态判断模块203，用于根据包的转发延时判断输出端口状态信息，以收到测量包发送模块102发送的测量包的时刻减去请求包发送模块202通告的时刻作为包的转发延时。因为交换网的最高优先级可以确保不发生拥塞，延时非常小，可以把由最高优先级发送的测量请求包当成瞬间达到。当输出端口状态信息处于正常状态且包的转发延时大于预设的拥塞发生判断域值时，状态判断模块203判断输出端口状态信息是拥塞状态；当输出端口状态信息处于拥塞状态且包的转发延时小于预设的拥塞解除判断域值时，状态判断模块203判断输出端口状态信息是正常状态。状态判断模块203判断成功后，使输出端口状态维护发布模块201更新输出端口状态信息。

以下结合图5、图6、图7和图8提供本发明的另一个实施例：

如图5所示，为本发明另一实施例的调度交换网数据包的方法流程图。该方法包括:

01、输出端口当前状态是否拥塞：如果拥塞则执行11，否则执行31;

11、接收测量请求包，当输出端口当前是拥塞状态时，接收交换网输出端口由所述交换网的最高优先级发送的测量请求包;

12、发送响应测量包，交换网输入端口收到测量请求包后由交换网的普通优先级发送响应测量包;

13、接收输出端口状态，接收所述交换网输出端口判断的输出端口状态;

在本发明实施例具体场景中，所述交换网输出端口判断输出端口状态是根据收到所述测量包的时刻减去所述测量请求包的发送时刻的差值作为所述包的转发延时判断的:

当所述包的转发延时小于预设的拥塞解除判断域值时，所述交换网输出端口判断它的状态改变为正常状态;

31、发送最高优先级测量包，当输出端口当前是正常状态时，由所述交换网的最高优先级向所述交换网输出端口发送最高优先级测量包;

32、发送普通测量包，由所述交换网的普通优先级向所述交换网输出端口发送普通测量包;

33、接收输出端口状态，接收所述交换网输出端口判断的输出端口状态;

在本发明实施例具体场景中，所述交换网输出端口判断输出端口状态是根据收到所述普通测量包的时刻减去收到所述最高优先级测量包的时刻的差值作为所述包的转发延时判断的:

当所述包的转发延时大于预设的拥塞发生判断域值时，所述交换网输出

端口判断它的状态改变为拥塞状态;

14、输出端口当前状态是否拥塞：在13或33接收输出端口状态之后根据接收的输出端口状态判断输出端口当前状态是否拥塞，如果拥塞则执行22，否则执行21;

21、不限流量，当所述输出端口的当前状态处于正常状态时，对所述交换网输出端口对应VoQ不做流量限制;

22、限制流量，当所述输出端口的当前状态处于拥塞状态时，对所述交换网输出端口对应VoQ的最大流量限制为交换网接口带宽除以交换网入口的总数。

如图6所示，为本发明另一实施例所述一种交换网数据包调度装置框图。该装置包括：VoQ状态维护模块101、测量包发送模块102和VoQ流量限制模块103，其中：

VoQ状态维护模块101，用于维护每个VoQ的状态信息。VoQ的状态有两种：正常状态和拥塞状态。VoQ状态维护模块101接收根据对包的转发延时进行测量的结果判断的输出端口状态，根据所述输出端口状态更新所述VoQ状态信息。当某个VoQ处于正常状态时，假如接收到对应输出端口的拥塞状态消息，则进入拥塞状态；否则保持正常状态。当某个VoQ处于拥塞状态时，假如接收到对应输出端口的正常状态消息，则进入正常状态；否则保持拥塞状态。

测量包发送模块102，主要用于执行两项任务：一是定时查询VoQ状态维护模块101的信息，为每个处于正常状态的VoQ发送测量包，这包括两个步骤，：先从交换网的最高优先级发最高优先级测量包，再从普通优先级发普通测量包；二是当测量包发送模块102接收到关于某个输出端口的请求测量包后，从普通优先级发一个到该端口的普通测量包。

VoQ流量限制模块，用于根据VoQ状态维护模块101的信息管理VoQ流量，对于所有处于正常状态的VoQ，不做流量限制。对于所有处于拥塞状态的VoQ，限制其最大流量为交换网接口带宽除以交换网入口的总数。

如图7所示，为本发明另一实施例所述交换网输出端口检控装置框图。该系统包括:

输出端口状态维护发布模块201，用于维护输出端口状态信息，输出端口的状态有两种：正常状态和拥塞状态。在输出端口的状态发生变化时（从正常状态变成拥塞状态，或者从拥塞状态变成正常状态），输出端口状态维护发布模块201向各个输入端口发布新的输出端口状态信息；或者输出端口状态维护发布模块201定期将当前输出端口状态信息发布给所有输入端口；或者输出端口状态维护发布模块201既在输出端口的状态发生变化时向各个输入端口发布新的输出端口状态信息，也定期将当前输出端口状态信息发布给所有输入端口。

请求包发送模块202，用于根据输出端口状态维护发布模块201的信息，当输出端口状态从正常状态变成拥塞状态时以及输出端口状态保持拥塞状态达到预设时间的倍数，通常是整数倍时，请求包发送模块202向输入端口发送测量请求包，并向状态判断模块203通告所述测量请求包的发送时刻。

状态判断模块203，用于根据包的转发延时判断输出端口状态信息，当输出端口状态信息处于正常状态时，状态判断模块203以收到输入端口发送的普通测量包的时刻减去收到最高优先级测量包的时刻的差值作为包的转发延时。当输出端口状态信息处于拥塞状态时，状态判断模块203以收到测量包的时刻减去请求包发送模块202通告的时刻作为包的转发延时。当输出端口状态信息处于正常状态且包的转发延时大于预设的拥塞发生判断域值时，状态判断模块203判断输出端口状态信息是拥塞状态；当输出端口状态信息处于拥塞状态且包的转发延时小于预设的拥塞解除判断域值时，状态判断模

块203判断输出端口状态信息是正常状态。状态判断模块203判断成功后，使输出端口状态维护发布模块201更新输出端口状态信息。

如图8所示，为本发明另一实施例所述一种调度交换网数据包的系统框图。该系统包括：输出端口状态维护发布模块201、请求包发送模块202和状态判断模块203，其中：

交换网输入端口管理装置1，用于发送测量包并管理VoQ;

交换网输出端口检控装置2，用于根据测量包的延时检测并发布输出端口的状态。

交换网输入端口管理装置1包括：VoQ状态维护模块101、测量包发送模块102和VoQ流量限制模块103，其中、

VoQ状态维护模块101，用于维护每个VoQ的状态信息。VoQ的状态有两种：正常状态和拥塞状态。VoQ状态维护模块101根据交换网输出端口检控装置2发布的输出端口状态信息维护VoQ的状态信息。当某个VoQ处于正常状态时，假如接收到对应输出端口的拥塞状态消息，则进入拥塞状态；否则保持正常状态。当某个VoQ处于拥塞状态时，假如接收到对应输出端口的正常状态消息，则进入正常状态；否则保持拥塞状态。

测量包发送模块102，主要用于执行两项任务：一是定时查询VoQ状态维护模块101的信息，为每个处于正常状态的VoQ发送测量包，包括：先从交换网的最高优先级发最高优先级测量包，再从普通优先级发普通测量包；二是当测量包发送模块102接收到交换网输出端口检控装置2的关于某个输出端口的请求测量包后，从普通优先级发一个到该端口的普通测量包。

VoQ流量限制模块103，用于根据VoQ状态维护模块101的信息管理VoQ流量，对于所有处于正常状态的VoQ，不做流量限制。对于所有处于拥塞状态的VoQ，限制其最大流量为交换网接口带宽除以交换网入口的总数。

交换网输出端口检控装置2包括：输出端口状态维护发布模块201、请求包发送模块202和状态判断模块203，其中：

输出端口状态维护发布模块201，用于维护输出端口状态信息，输出端口的状态有两种：正常状态和拥塞状态。在输出端口的状态发生变化时（从正常状态变成拥塞状态，或者从拥塞状态变成正常状态），输出端口状态维护发布模块201向各个VoQ状态维护模块101发布新的输出端口状态信息；或者输出端口状态维护发布模块201定期将当前输出端口状态信息发布给所有入口的VoQ状态维护模块101；或者输出端口状态维护发布模块201既在输出端口的状态发生变化时向各个VoQ状态维护模块101发布新的输出端口状态信息，也定期将当前输出端口状态信息发布给所有入口的VoQ状态维护模块101。

请求包发送模块202，用于根据输出端口状态维护发布模块201的信息，当输出端口状态从正常状态变成拥塞状态时以及输出端口状态保持拥塞状态达到预设时间的倍数，通常是整数倍时，请求包发送模块202向交换网输入端口管理装置1发送测量请求包，并向状态判断模块203通告所述测量请求包的发送时刻。

状态判断模块203，用于根据包的转发延时判断输出端口状态信息，当输出端口状态信息处于正常状态时，状态判断模块203以收到测量包发送模块102发送的普通测量包的时刻减去收到最高优先级测量包的时刻的差值作为包的转发延时。因为交换网的最高优先级可以确保不发生拥塞，延时非常小，可以把由最高优先级发送的最高优先级测量包当成瞬间达到；当输出端口状态信息处于拥塞状态时，状态判断模块203以收到测量包的时刻减去请求包发送模块202通告的时刻作为包的转发延时。因为交换网的最高优先级可以确保不发生拥塞，延时非常小，可以把由最高优先级发送的测量请求包当成瞬间达到。当输出端口状态信息处于正常状态且包的转发延时大于预设

的拥塞发生判断域值时，状态判断模块203判断输出端口状态信息是拥塞状态；当输出端口状态信息处于拥塞状态且包的转发延时小于预设的拥塞解除判断域值时，状态判断模块203判断输出端口状态信息是正常状态。状态判断模块203判断成功后，使输出端口状态维护发布模块201更新输出端口状态信息。

以下结合图9、图10、图11和图12提供本发明的一个实施例:

如图9所示，为本发明另一实施例的调度交换网数据包的方法流程图。在所述交换网支持全局时钟同步时，不仅本发明的前两个实施例，以下实施例也可以提供调度交换网数据包的方法。该方法包括:

01、输出端口当前状态是否拥塞：如果拥塞则执行11，否则执行41;

11、接收测量请求包，当输出端口当前是拥塞状态时，接收交换网输出端口由所述交换网的最高优先级发送的测量请求包;

12、发送响应测量包，交换网输入端口收到测量请求包后由交换网的普通优先级发送响应测量包;

13、接收输出端口状态，接收所述交换网输出端口判断的输出端口状态;

在本发明实施例具体场景中，所述交换网输出端口判断输出端口状态是根据收到所述测量包的时刻减去所述测量请求包的发送时刻的差值作为所述包的转发延时判断的:

当所述包的转发延时小于预设的拥塞解除判断域值时，所述交换网输出端口判断它的状态改变为正常状态;

41、发送时戳测量包，当输出端口当前是正常状态时，向所述交换网输出端口发送打上进入交换网时戳的测量包;

42、接收输出端口状态，接收所述交换网输出端口判断的输出端口状

态;

在本发明实施例具体场景中，所述交换网输出端口判断输出端口状态是根据收到所述打上交换网时戳的测量包的本地时刻减去所述测量包的进入交换网时戳作为所述包的转发延时判断的:

当所述包的转发延时大于预设的拥塞发生判断域值时，所述交换网输出端口判断它的状态改变为拥塞状态;

14、输出端口当前状态是否拥塞：在13或42接收输出端口状态之后根据接收的输出端口状态判断输出端口当前状态是否拥塞，如果拥塞则执行22，否则执行21;

21、不限流量，当所述输出端口的当前状态处于正常状态时，对所述交换网输出端口对应VoQ不做流量限制;

22、限制流量，当所述输出端口的当前状态处于拥塞状态时，对所述交换网输出端口对应VoQ的最大流量限制为交换网接口带宽除以交换网入口的总数。

如图10所示，为本发明另一实施例所述在交换网支持全局时钟同步时，一种交换网数据包调度装置框图。在所述交换网支持全局时钟同步时，不仅本发明的前两个实施例，该实施例也可以提供交换网数据包调度装置。该装置包括：VoQ状态维护模块101、测量包发送模块102和VoQ流量限制模块103，其中:

VoQ状态维护模块101，用于维护每个VoQ的状态信息。VoQ的状态有两种：正常状态和拥塞状态。VoQ状态维护模块101接收根据对包的转发延时进行测量的结果判断的输出端口状态，根据所述输出端口状态更新所述VoQ状态信息。当某个VoQ处于正常状态时，假如接收到对应输出端口的拥塞状态消息，则进入拥塞状态；否则保持正常状态。当某个VoQ处于拥塞状

态时，假如接收到对应输出端口的正常状态消息，则进入正常状态；否则保持拥塞状态。

测量包发送模块102，主要用于执行两项任务：一是定时查询VoQ状态维护模块101的信息，为每个处于正常状态的VoQ发送时戳测量包，该时戳测量包打上进入交换网的时戳；二是当测量包发送模块102接收到关于某个输出端口的请求测量包后，发一个到该端口响应测量包。

VoQ流量限制模块103，用于根据VoQ状态维护模块101的信息管理VoQ流量，对于所有处于正常状态的VoQ，不做流量限制。对于所有处于拥塞状态的VoQ，限制其最大流量为交换网接口带宽除以交换网入口的总数。

如图12所示，为本发明另一实施例所述在交换网支持全局时钟同步时，一种交换网输出端口检控装置框图。在所述交换网支持全局时钟同步时，不仅本发明的前两个实施例，该实施例也可以提供交换网输出端口检控装置。该装置包括：输出端口状态维护发布模块201、请求包发送模块202和状态判断模块203，其中：

输出端口状态维护发布模块201，用于维护输出端口状态信息，输出端口的状态有两种：正常状态和拥塞状态。在输出端口的状态发生变化时（从正常状态变成拥塞状态，或者从拥塞状态变成正常状态），输出端口状态维护发布模块201向各个输入端口发布新的输出端口状态信息；或者输出端口状态维护发布模块201定期将当前输出端口状态信息发布给所有输入端口；或者输出端口状态维护发布模块201既在输出端口的状态发生变化时向各个输入端口发布新的输出端口状态信息，也定期将当前输出端口状态信息发布给所有输入端口。

请求包发送模块202，用于查询输出端口状态维护发布模块201的信息，当输出端口状态从正常状态变成拥塞状态时以及输出端口状态保持拥塞状态达到预设时间的倍数，通常为整数倍时，请求包发送模块202向交换网输入

端口发送测量请求包，并向状态判断模块203通告所述测量请求包的发送时刻。

状态判断模块203，用于根据包的转发延时判断输出端口状态信息，当输出端口状态信息处于正常状态时，状态判断模块203以收到的时戳测量包的时戳减去收到该时戳测量包的时刻的差值作为包的转发延时；当输出端口状态信息处于拥塞状态时，状态判断模块203以收到响应测量包的时刻减去请求包发送模块202通告的时刻作为包的转发延时。当输出端口状态信息处于正常状态且包的转发延时大于预设的拥塞发生判断域值时，状态判断模块203判断输出端口状态信息是拥塞状态；当输出端口状态信息处于拥塞状态且包的转发延时小于预设的拥塞解除判断域值时，状态判断模块203判断输出端口状态信息是正常状态。状态判断模块203判断成功后，使输出端口状态维护发布模块201更新输出端口状态信息。

如图12所示，为本发明另一实施例所述在交换网支持全局时钟同步时，一种调度交换网数据包的系统框图。在所述交换网支持全局时钟同步时，不仅本发明的前两个实施例，该实施例也可以提供调度交换网数据包的系统。该系统包括:

交换网输入端口管理装置1，用于发送测量包并管理VoQ;

交换网输出端口检控装置2，用于根据测量包的延时检测并发布输出端口的状态。

交换网输入端口管理装置1包括：VoQ状态维护模块101、测量包发送模块102和VoQ流量限制模块103，其中:

VoQ状态维护模块101，用于维护每个VoQ的状态信息。VoQ的状态有两种：正常状态和拥塞状态。VoQ状态维护模块101根据交换网输出端口检控装置2发布的输出端口状态信息维护VoQ的状态信息。当某个VoQ处于正常状态时，假如接收到对应输出端口的拥塞状态消息，则进入拥塞状态；否则

保持正常状态。当某个VoQ处于拥塞状态时，假如接收到对应输出端口的正常状态消息，则进入正常状态；否则保持拥塞状态。

测量包发送模块102，主要用于执行两项任务：一是定时查询VoQ状态维护模块101的信息，为每个处于正常状态的VoQ发送时戳测量包，该时戳测量包打上进入交换网的时戳；二是当测量包发送模块102接收到交换网输出端口检控装置2的关于某个输出端口的请求测量包后，发一个到该端口响应测量包。

VoQ流量限制模块103，用于根据VoQ状态维护模块101的信息管理VoQ流量，对于所有处于正常状态的VoQ，不做流量限制。对于所有处于拥塞状态的VoQ，限制其最大流量为交换网接口带宽除以交换网入口的总数。

交换网输出端口检控装置2包括：输出端口状态维护发布模块201、请求包发送模块202和状态判断模块203，其中：

输出端口状态维护发布模块201，用于维护输出端口状态信息，输出端口的状态有两种：正常状态和拥塞状态。在输出端口的状态发生变化时（从正常状态变成拥塞状态，或者从拥塞状态变成正常状态），输出端口状态维护发布模块201向各个VoQ状态维护模块101发布新的输出端口状态信息；或者输出端口状态维护发布模块201定期将当前输出端口状态信息发布给所有入口的VoQ状态维护模块101；或者输出端口状态维护发布模块201既在输出端口的状态发生变化时向各个VoQ状态维护模块101发布新的输出端口状态信息，也定期将当前输出端口状态信息发布给所有入口的VoQ状态维护模块101。

请求包发送模块202，用于查询输出端口状态维护发布模块201的信息，当输出端口状态从正常状态变成拥塞状态时以及输出端口状态保持拥塞状态达到预设时间的倍数，通常为整数倍时，请求包发送模块202向交换网输入端口管理装置1发送测量请求包，并向状态判断模块203通告所述测量请求包

的发送时刻。

状态判断模块203，用于根据包的转发延时判断输出端口状态信息，当输出端口状态信息处于正常状态时，状态判断模块203以收到的时戳测量包的时戳减去收到该时戳测量包的时刻的差值作为包的转发延时；当输出端口状态信息处于拥塞状态时，状态判断模块203以收到响应测量包的时刻减去请求包发送模块202通告的时刻作为包的转发延时。因为交换网的最高优先级可以确保不发生拥塞，延时非常小，可以把由最高优先级发送的测量请求包当成瞬间达到。当输出端口状态信息处于正常状态且包的转发延时大于预设的拥塞发生判断域值时，状态判断模块203判断输出端口状态信息是拥塞状态；当输出端口状态信息处于拥塞状态且包的转发延时小于预设的拥塞解除判断域值时，状态判断模块203判断输出端口状态信息是正常状态。状态判断模块203判断成功后，使输出端口状态维护发布模块201更新输出端口状态信息。

以下结合图13提供本发明的一个实施例场景：

包的转发延时与设备的缓存使用情况是相关的，通过测量一个包的转发延时，可以计算出设备在转发这个包之前的缓存使用数量。在流量比较稳定（变化周期与测量间隔相比很大，10倍左右），则还可以粗略估算出当前的缓存使用数量。

当交换网的某个输出端口发生拥塞时，它所使用的缓存数量不断增加，从该端口转发的包的延时也将不断加大。反之拥塞解除后包的转发延时会不断减小。因此通过延时测量，在交换网出口可以估计出交换网是否发生了拥塞，以及拥塞是否已经解除。

发生拥塞的输出端口再将这个信息反馈到对应的输入端口，即可用作VoQ调度器的控制信息。

本发明实施例通过发送测量包检测包的转发延时，无需交换网支持传递

反压信息就能判断输出端口是否拥塞，提供了在交换网不支持传递反压信息、或者交换网和VoQ调度器消息格式不兼容的情况下，解决队头拥塞的方案，从而降低了对交换网兼容性的要求。设备商可以灵活选择不同的交换网搭建系统而无须顾虑不同厂家交换网消息格式的兼容性问题，也可以使用LSW芯片来作为交换网片，从而有利于设备商降低设备成本。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分步骤是可以通过程序来指令相关的转发平面完成，所述的程序可以存储于计算机可读存储介质中，所述存储介质可以是ROM/RAM，磁盘或光盘等。

以上所述，仅为本发明较佳的具体实施方式，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，可轻易想到的变化或替换，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该以权利要求的保护范围为准。

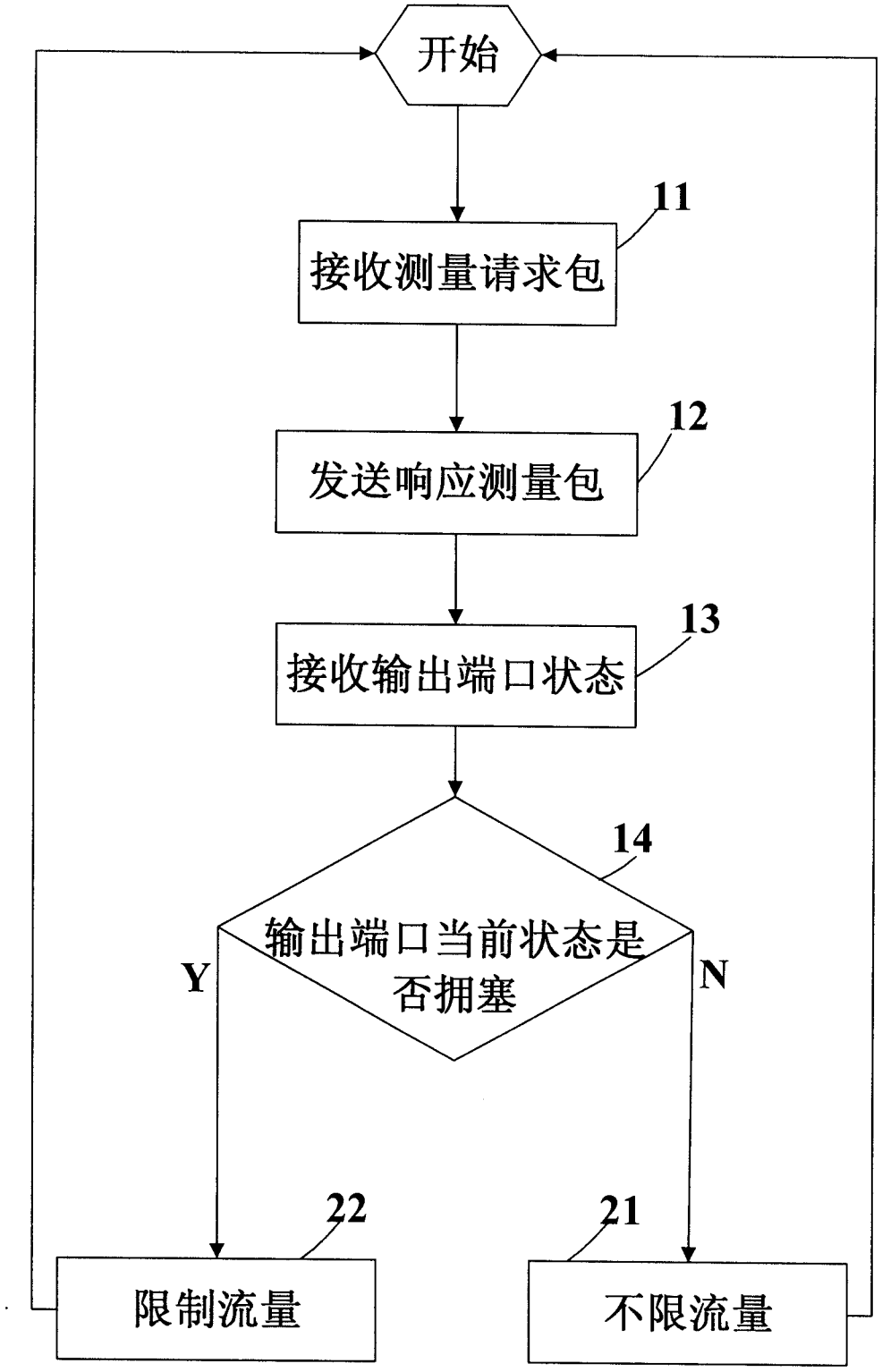

开始
11
接收测量请求包
12
发送响应测量包
13
接收输出端口状态
14
输出端口当前状态是否拥塞
Y
N
22
限制流量
21
不限流量

图1

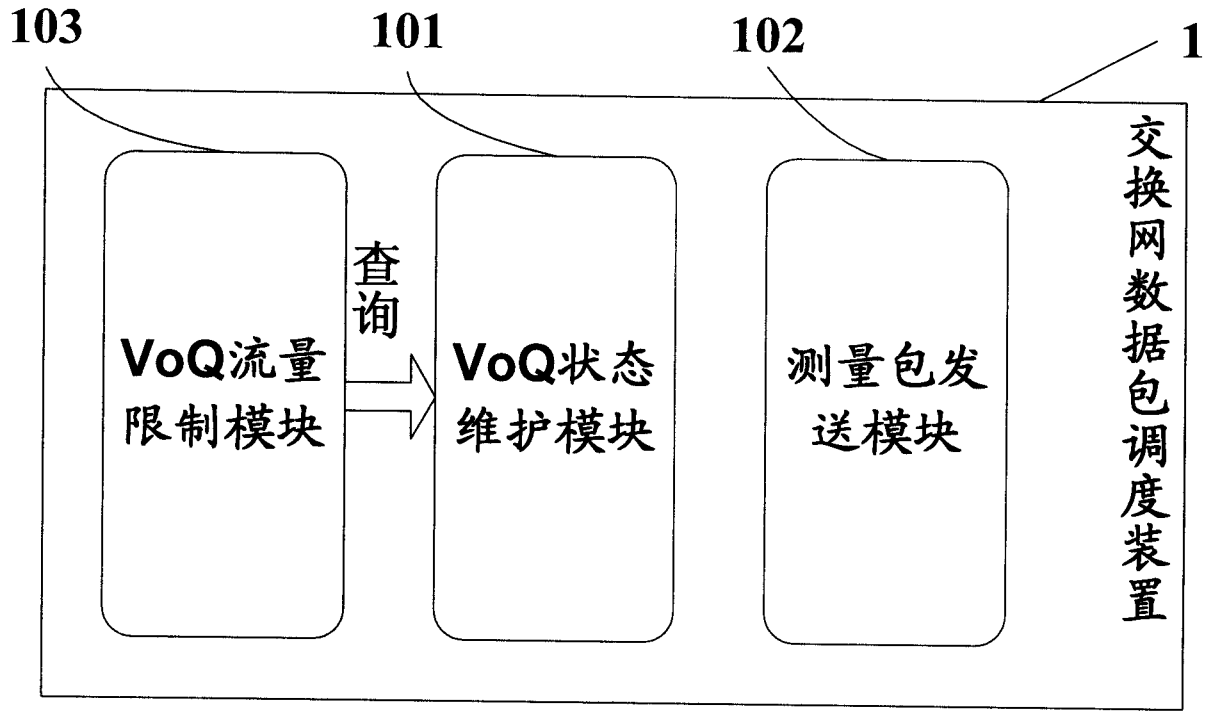

103
101
102
1
交换网数据包调度装置
VoQ流量限制模块
查询
VoQ状态维护模块
测量包发送模块

图2

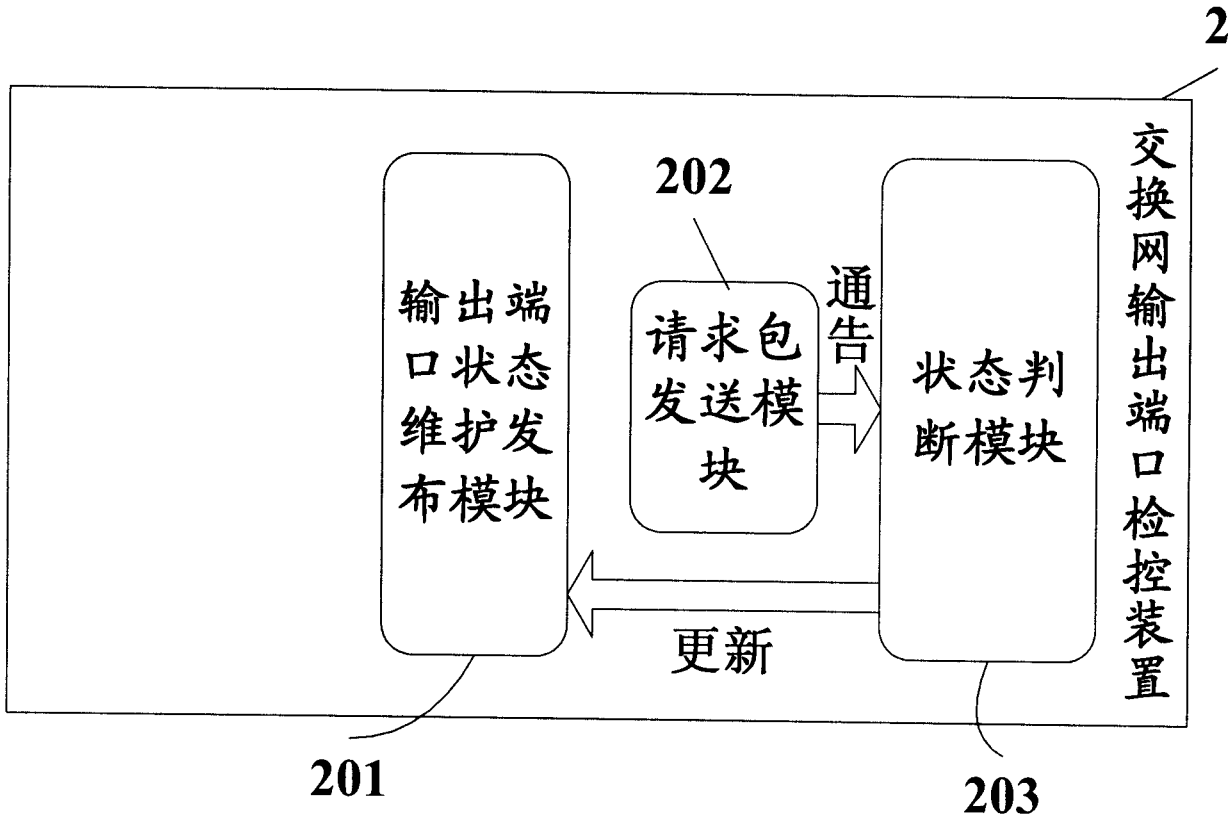

2
交换网输出端口检控装置
202
端口状态发送模块
请求包发送模块
通告
状态判断模块
更新
201
203

图3

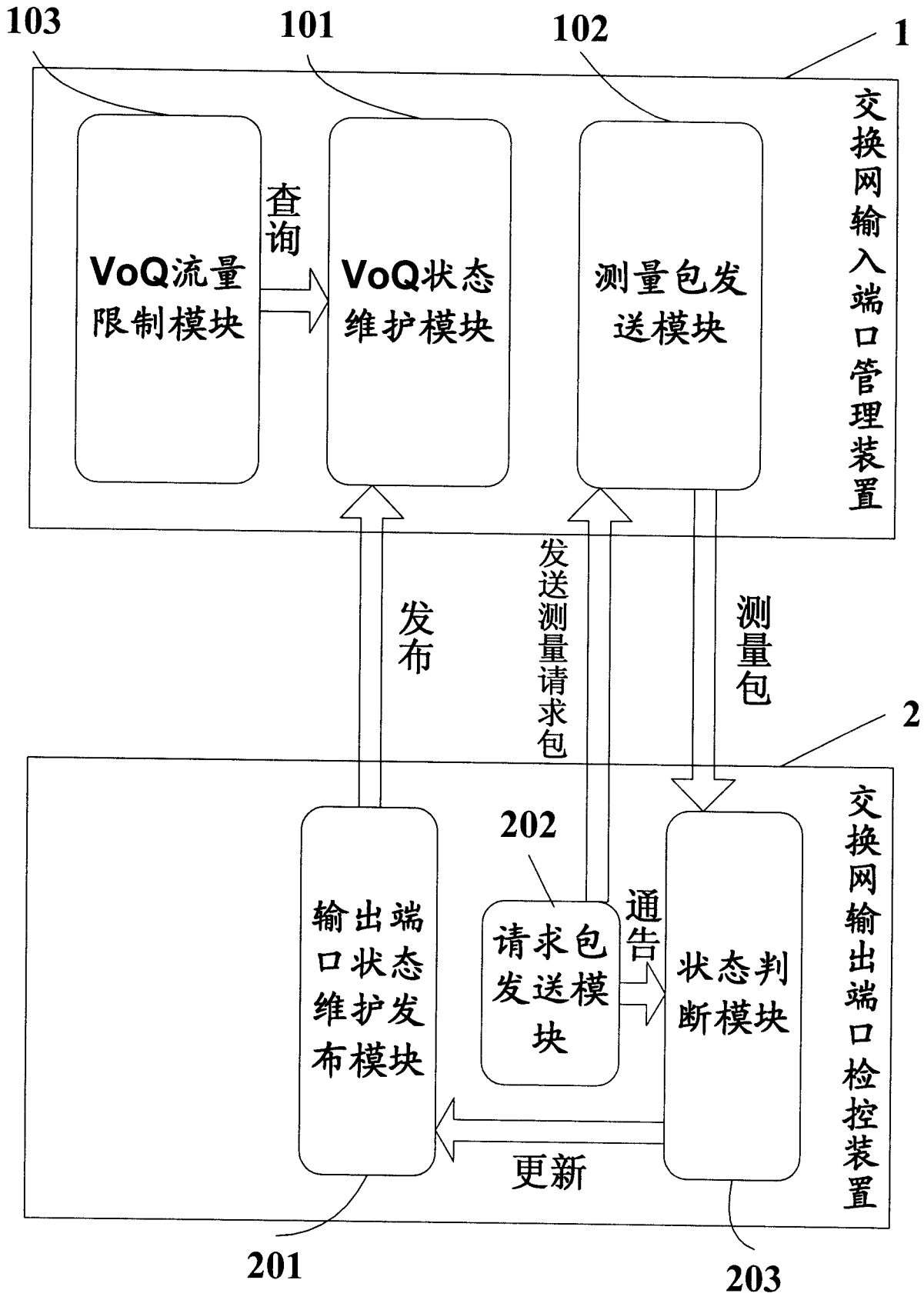

103
101
102
1
交换网输入端口管理装置
VoQ流量限制模块
查询
VoQ状态维护模块
测量包发送模块
发布
发送测量请求包
测量包
2
交换网输出端口检控装置
202
输出端口状态维护发布模块
请求包发送模块
通告
状态判断模块
更新
201
203

图4

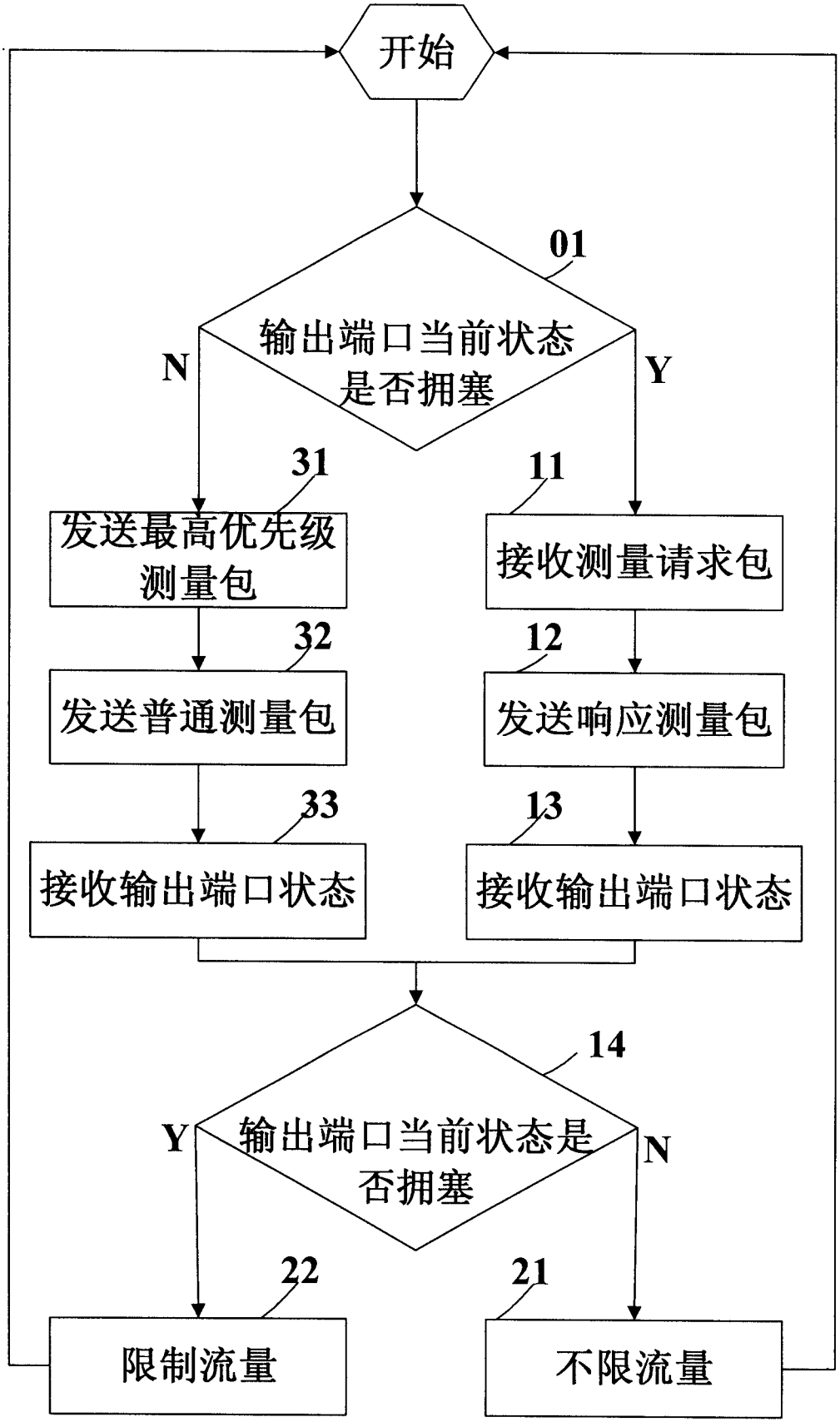

开始
01
输出端口当前状态是否拥塞
N
Y
31
发送最高优先级测量包
11
接收测量请求包
32
发送普通测量包
12
发送响应测量包
33
接收输出端口状态
13
接收输出端口状态
14
输出端口当前状态是否拥塞
Y
N
22
限制流量
21
不限流量

图5

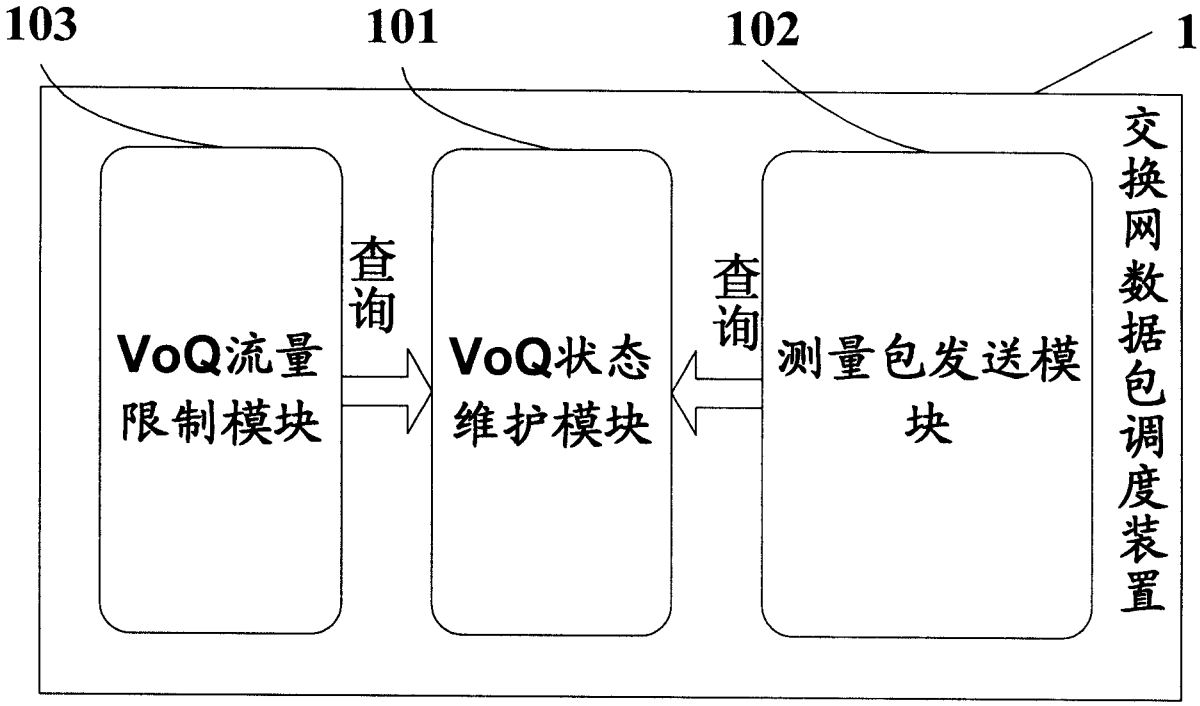

103
101
102
1
交换网数据包调度装置
VoQ流量限制模块
查询
VoQ状态维护模块
查询
测量包发送模块

图6

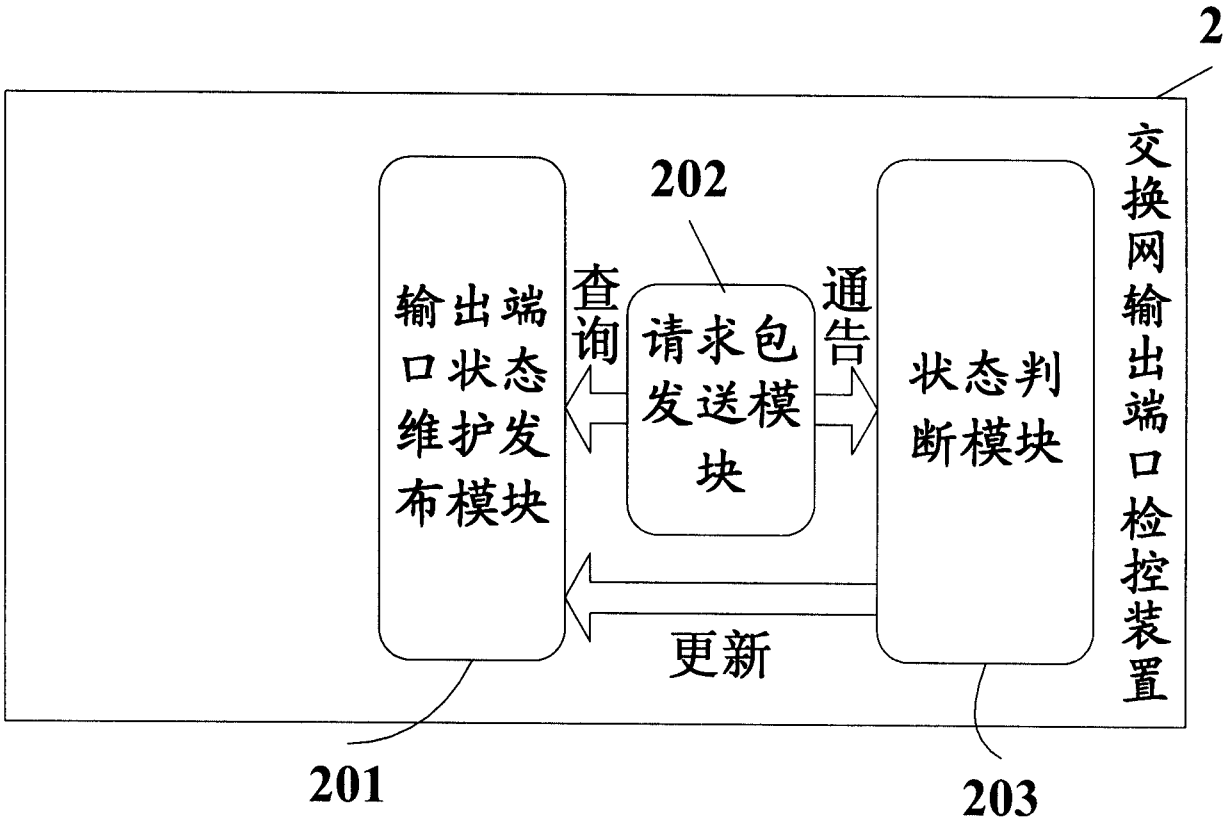

2
交换网输出端口检控装置
202
端口状态发送模块
查询
请求包发送模块
通告
状态判断模块
更新
201
203

图7

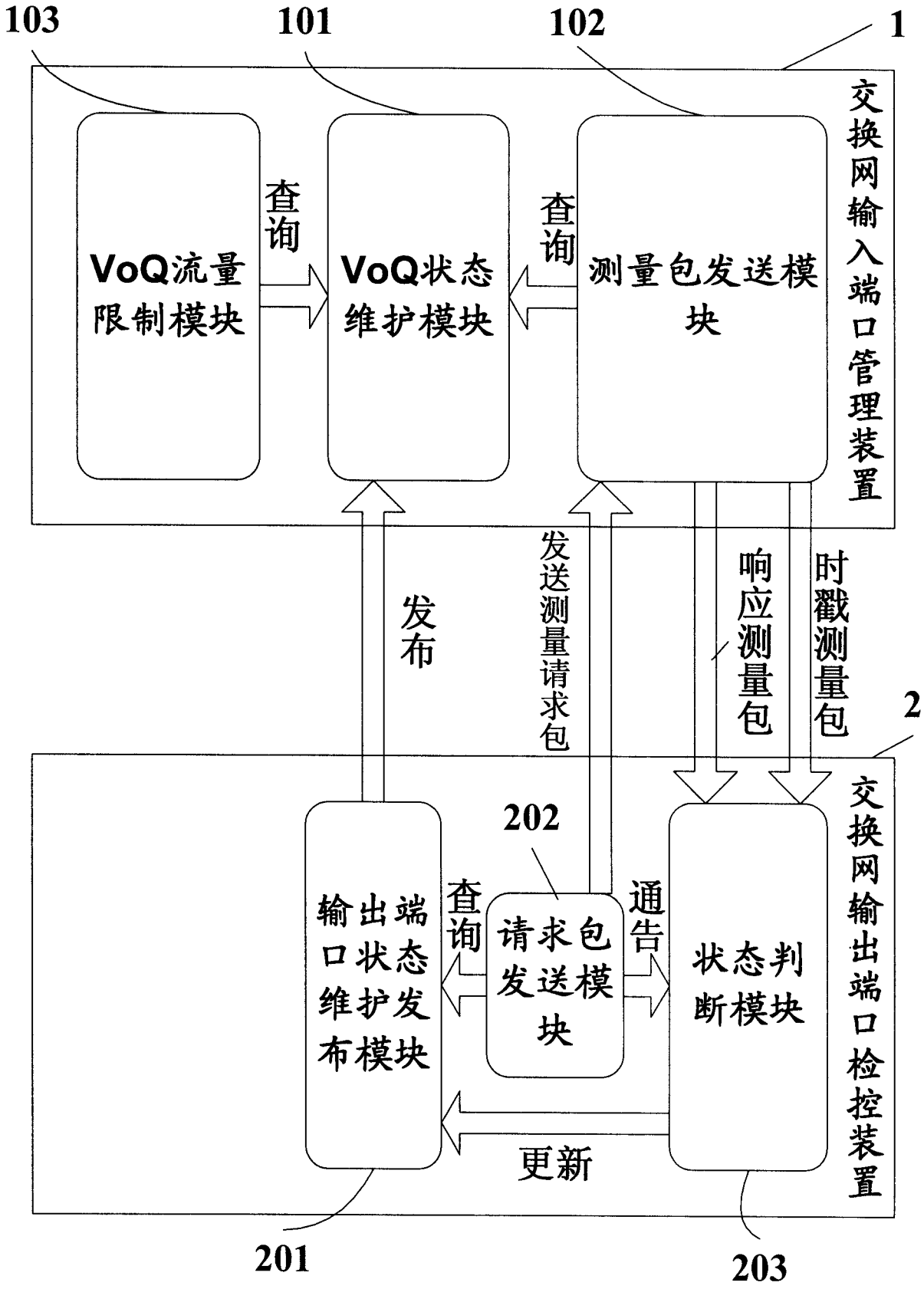

103
101
102
1
交换网输入端口管理装置
VoQ流量限制模块
查询
VoQ状态维护模块
查询
测量包发送模块
发布
发送测量请求包
响应测量包
时戳测量包
2
交换网输出端口检控装置
202
输出端口状态维护发布模块
查询
请求包发送模块
通告
状态判断模块
更新
201
203

图8

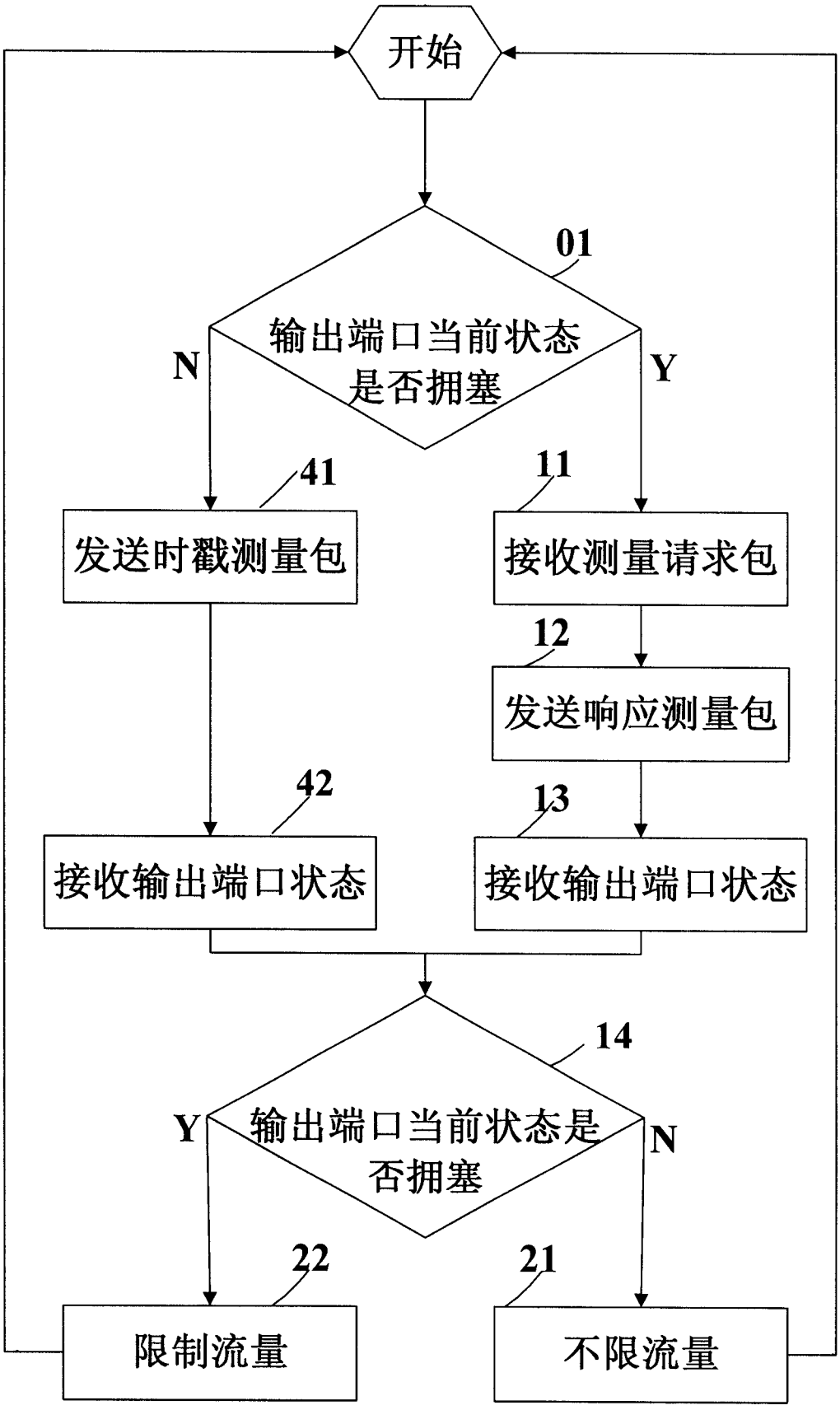

开始
01
输出端口当前状态是否拥塞
N
Y
41
发送时戳测量包
11
接收测量请求包
12
发送响应测量包
42
接收输出端口状态
13
接收输出端口状态
14
输出端口当前状态是否拥塞
Y
N
22
限制流量
21
不限流量

图9

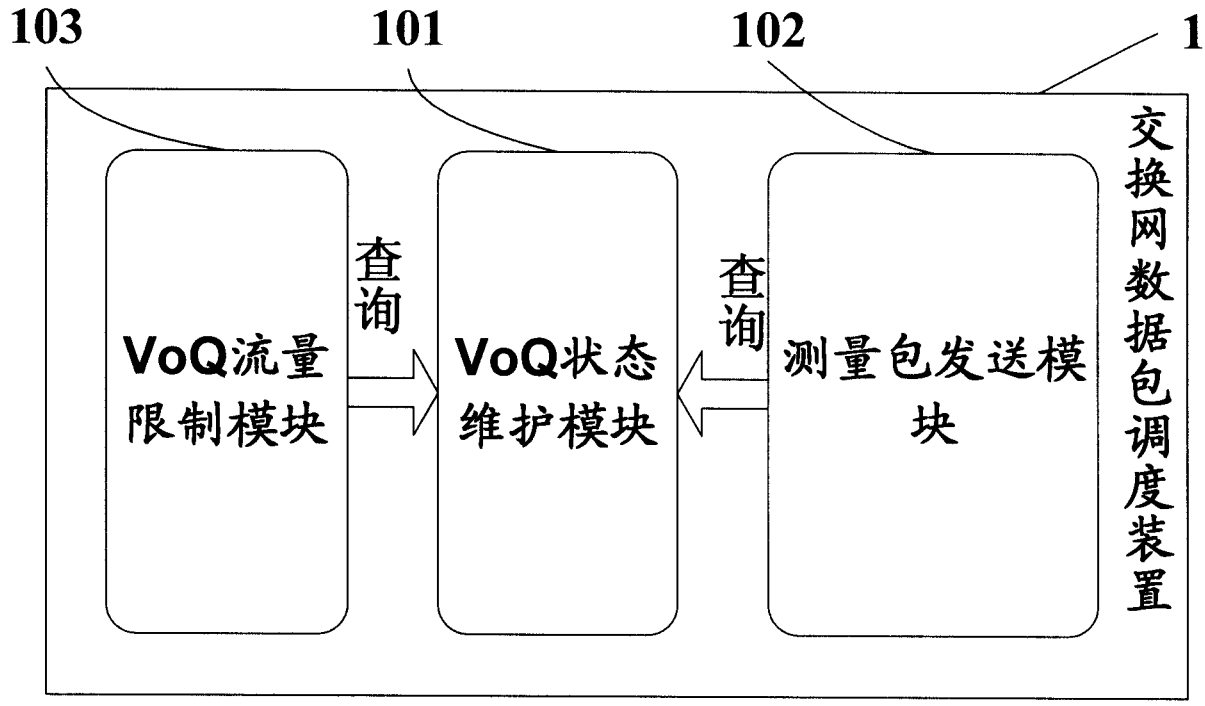

103
101
102
1
交换网数据包调度装置
VoQ流量限制模块
查询
VoQ状态维护模块
查询
测量包发送模块

图10

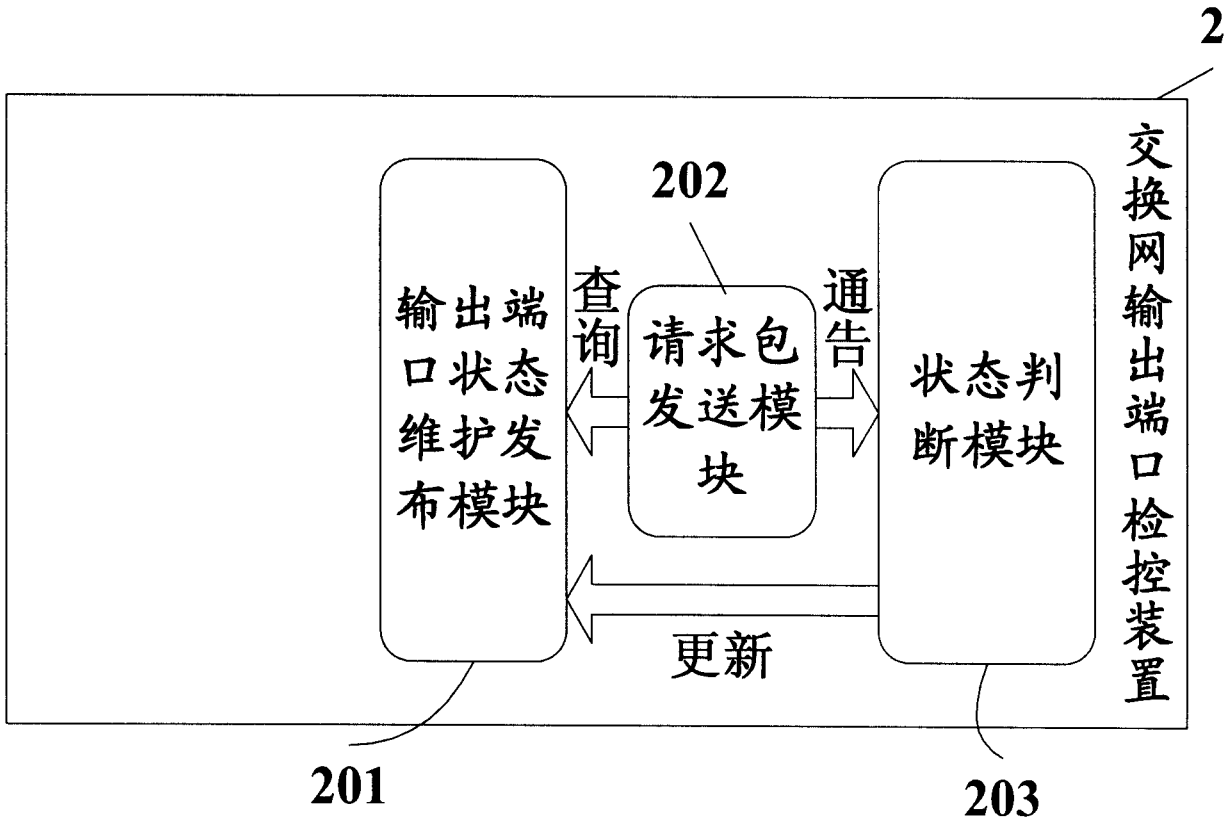

2
交换网输出端口检控装置
202
输出端口状态维护发布模块
查询
请求包发送模块
通告
状态判断模块
更新
201
203

图11

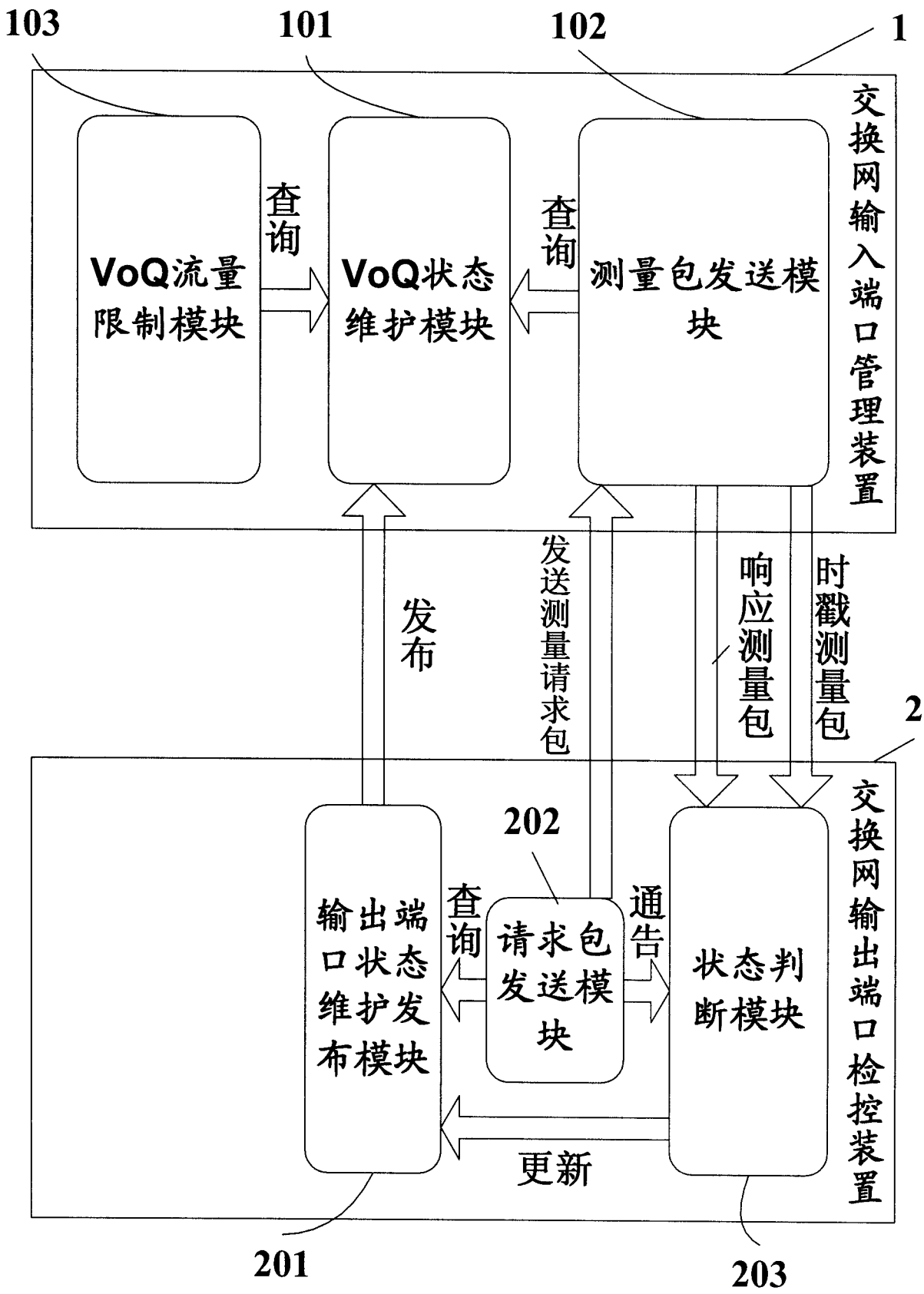

103
101
102
1
交换网输入端口管理装置
VoQ流量限制模块
查询
VoQ状态维护模块
查询
测量包发送模块
发布
发送测量请求包
响应测量包
时戳测量包
2
交换网输出端口检控装置
202
输出端口状态维护模块
查询
请求包发送模块
通告
状态判断模块
更新
201
203

图12

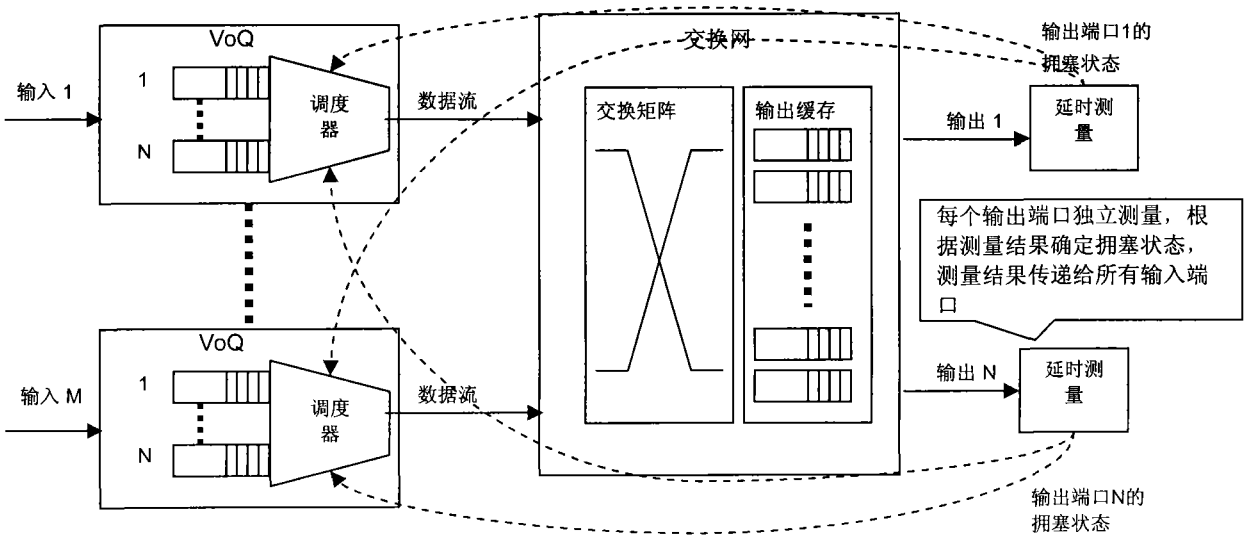

VoQ
1
N
输入 1
调度器
数据流
交换网
交换矩阵
输出缓存
输出端口1的拥塞状态
输出 1
延时测量
每个输出端口独立测量，根据测量结果确定拥塞状态，测量结果传递给所有输入端口
VoQ
1
N
输入 M
调度器
数据流
输出 N
延时测量
输出端口N的拥塞状态

图13