[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04L 12/56 (2006.01)
H04L 12/46 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200810133299.2

[43] 公开日 2008 年 12 月 17 日

[11] 公开号 CN 101325557A

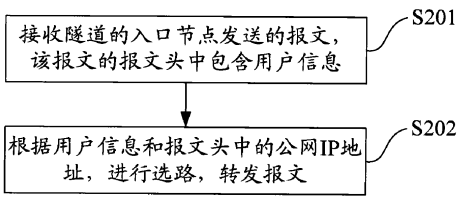
[22] 申请日 2008.7.25
[21] 申请号 200810133299.2
[71] 申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
[72] 发明人 田小辉
[74] 专利代理机构 北京挺立专利事务所
代理人 叶树明
权利要求书 2 页 说明书 9 页 附图 4 页

[54] 发明名称

一种隧道负载分担的方法、系统和装置

[57] 摘要

本发明实施例公开了一种隧道负载分担的方法、系统和装置，所述隧道负载分担的方法包括：接收隧道的入口节点发送的报文，所述报文的报文头中包含用户信息；根据所述用户信息和所述报文头中的公网因特网协议 IP 地址，进行选路，转发所述报文。 本发明实施例实现了在隧道数目有限或隧道地址分散时，负载分担节点也可以将流量均匀地分布到每条链路上。 并且本发明实施例在对报文进行分片时，将包含用户信息的 Options 域复制到分片后每个报文的报文头中，从而避免了报文乱序问题的产生。



1、一种隧道负载分担的方法，其特征在于，包括:

接收隧道的入口节点发送的报文，所述报文的报文头中包含用户信息;

根据所述用户信息和所述报文头中的公网因特网协议 IP 地址，进行选路，转发所述报文。

2、如权利要求 1 所述隧道负载分担的方法，其特征在于，所述用户信息由所述隧道的入口节点根据用户的 IP 地址、媒体接入控制 MAC 地址和会话标识中的一种或几种，进行哈希获取，并且所述用户信息封装在所述报文头的选项 Options 域中。

3、如权利要求 2 所述隧道负载分担的方法，其特征在于，在所述根据所述用户信息和所述报文头中的公网 IP 地址，进行选路，转发所述报文之前，还包括:

在对所述隧道的入口节点发送的报文进行分片时，将所述 Options 域复制到分片后每个报文的报文头中。

4、如权利要求 2 所述隧道负载分担的方法，其特征在于，所述 Options 域包括选项类型字段、选项长度字段和选项数据字段，所述选项数据字段承载所述用户信息。

5、一种隧道负载分担的系统，其特征在于，包括:

隧道的入口节点，用于获取用户的用户信息，将所述用户信息封装在报文头中，并发送携带所述报文头的报文;

负载分担节点，用于接收所述隧道的入口节点发送的报文，根据所述报文头中的用户信息和公网 IP 地址，进行选路，转发所述报文。

6、如权利要求 5 所述负载分担的系统，其特征在于，还包括:

转发节点，用于透传所述负载分担节点转发的报文;

隧道的出口节点，用于接收所述转发节点透传的报文，解析所述报文，当所述报文为出隧道报文时，对所述报文进行解封装，去掉所述报文的报文头，并对所述报文进行出隧道处理。

7、一种隧道的入口节点，其特征在于，包括:

获取模块，用于获取用户的用户信息;

封装模块，用于将所述获取模块获取的用户信息封装在报文头中；

发送模块，用于发送携带所述封装模块封装的报文头的报文。

8、如权利要求 7 所述隧道的入口节点，其特征在于，所述获取模块包括：

哈希子模块，用于根据用户的 IP 地址、媒体接入控制 MAC 地址和会话标识中的一种或几种，进行哈希获取所述用户的用户信息。

9、一种负载分担节点，其特征在于，包括：

接收模块，用于接收隧道的入口节点发送的报文，所述报文的报文头中包含用户信息；

报文转发模块，用于根据所述接收模块接收的报文头中的用户信息和公网 IP 地址，进行选路，转发所述报文。

10、如权利要求 9 所述负载分担节点，其特征在于，还包括：

复制模块，用于在对所述接收模块接收的报文进行分片时，将所述接收模块接收的报文头中的 Options 域复制到分片后每个报文的报文头中，所述 Options 域携带所述用户信息。

一种隧道负载分担的方法、系统和装置

技术领域

本发明实施例涉及通信技术领域，特别涉及一种隧道负载分担的方法、系统和装置。

背景技术

在网络中的某个节点上，路由转发时可能存在 2 条以上的等价链路，为了保证每条链路的流量均匀，通常需要按照某种方法将流量均匀地分布在每条链路上。

现有技术通常使用的是 HASH（hash，哈希）算法。此方法就是根据流量的属性，例如：IPv4（Internet Protocol version 4，因特网协议版本 4）报文的 IP 地址、二层以太报文的 MAC（Media Access Control，媒体接入控制）地址等可以标识某一条流量的参数，将不同属性的流量散列到不同的链路上去。而且，在大多数情况下，流量在传输中是需要保证顺序的，因此在 HASH 的过程中还要保证同一条流一定要选择同一个链路，以免乱序。

隧道技术是使用一种协议封装另外一种协议报文的技术，封装协议本身也可以被其他封装协议所封装或承载。对用户来说，隧道是其 PSTN（Public Switched Telephone Network，公共交换电话网）/ISDN（Integrated Services Digital Network，综合业务数字网）链路的逻辑延伸，在使用上与实际物理链路相同。

基于 IPv4 网络的隧道是承载在 IPv4 网络中的隧道技术，在这种技术中，用户报文都被封装到 IPv4 协议内部进行传输。

现有技术中，网络中的节点收到一个 IPv4 报文，在有多条转发路径、同时需要保证报文顺序时，通常使用 IPv4 头中的源 IP（Internet Protocol，因特网协议）和目的 IP 标识一条数据流。因此，对于承载在 IPv4 网络的隧道中的各种用户报文而言，在网络侧只会使用隧道的公网 IPv4 地址进行 HASH。

采用上述方法，如果由于网络布局的原因，导致隧道数目较少或隧道地址不连续，则在IPv4网络侧，遇到需要负载分担的情形时，会导致HASH的结果不均匀，将所有用户或大部分用户的流量都选择到一条或少数几条链路上，而其他链路仍然空闲。在用户流量大的情况下，被选择的链路容易产生拥塞，发生丢包现象。

现有技术提出了一种隧道负载分担的方法，可以在负载分担节点（例如：IPv4 核心节点）上解析内部隧道协议报文，再根据内部用户的 IP 地址进行HASH，选择转发的链路。例如：对于L2TP（Layer Two Tunneling Protocol，二层隧道协议）隧道，负载分担节点需要解析 IPv4 报文中的 UDP（User Datagram Protocol，用户数据报协议）端口号、L2TP头以及内部的私网IPv4地址等信息，再根据内部的私网IPv4地址进行HASH，选择转发的链路。

但是，上述隧道负载分担的方法，判断内部隧道协议的分支比较复杂，比如，一个用户报文内部可以封装L2TP、GRE（Generic Routing Encapsulation，通用路由封装）、IPSec（Internet Protocol Security extensions，IP协议安全扩展）等多种隧道的用户报文。如果负载分担节点对报文中封装的所有报文全部进行解析，算法复杂度很大，并且在解析内部协议时，需要从外部存储器中读取并进行分析的报文长度很大，导致公网侧的核心路由器转发性能受到很大影响。

如果公网侧的IPv4报文被分片，使用此方案则会导致报文的乱序（接收到报文的顺序与发送的顺序不符）。以 GRE 报文为例，如图 1 所示。不分片的报文以及分片报文的首片可以根据用户内部协议报文进行哈希；而由于后续片中没有了 GRE 头，无法根据内部协议做 HASH，只能根据公网侧 IPv4 地址进行 HASH，有可能和首片选择不同的转发路径。如果后续片比较晚到达，在重组侧就会导致分片报文的重组被延误，后面的不分片报文提前到达，导致报文产生乱序。

发明内容

为了更清楚地说明本发明实施例的技术方案，下面将对实施例描述中所

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。

本发明实施例提供一种隧道负载分担的方法，以实现隧道数目有限或隧道地址分散时，将流量均匀地分布到每条链路上。

为达到上述目的，本发明实施例一方面提供一种隧道负载分担的方法，包括:

接收隧道的入口节点发送的报文，所述报文的报文头中包含用户信息;

根据所述用户信息和所述报文头中的公网因特网协议 IP 地址，进行选路，转发所述报文。

另一方面，本发明实施例还提供一种隧道负载分担的系统，包括:

隧道的入口节点，用于获取用户的用户信息，将所述用户信息封装在报文头中，并发送携带所述报文头的报文;

负载分担节点，用于接收所述隧道的入口节点发送的报文，根据所述报文头中的用户信息和公网 IP 地址，进行选路，转发所述报文。

再一方面，本发明实施例还提供一种隧道的入口节点，包括:

获取模块，用于获取用户的用户信息;

封装模块，用于将所述获取模块获取的用户信息封装在报文头中;

发送模块，用于发送携带所述封装模块封装的报文头的报文

再一方面，本发明实施例还提供一种负载分担节点，包括:

接收模块，用于接收隧道的入口节点发送的报文，所述报文的报文头中包含用户信息;

报文转发模块，用于根据所述接收模块接收的报文头中的用户信息和公网 IP 地址，进行选路，转发所述报文。

与现有技术相比，本发明实施例具有以下优点：通过本发明实施例，负载分担节点接收包含用户信息的报文，并根据报文头中的 IP 地址和用户信息进行选路，转发该报文。因此，在隧道数目有限或隧道地址分散时，负载分担节点也可以将流量均匀地分布到每条链路上。

附图说明

图 1 为现有技术产生报文乱序问题的示意图;

图 2 为本发明实施例隧道负载分担的方法的流程图;

图 3 为本发明实施例隧道负载分担的方法的组网示意图;

图 4 为本发明实施例隧道的入口节点的操作流程示意图;

图 5 为本发明实施例 IPv4 报文头的结构图;

图 6 为本发明实施例 Options 域的结构图;

图 7 为本发明实施例隧道负载分担的系统的结构图;

图 8 为本发明实施例隧道的入口节点的结构图;

图 9 为本发明实施例负载分担节点的结构图。

具体实施方式

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的一部分实施例，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。

本发明实施例提供了一种隧道负载分担的方法，将用户信息带到公网侧，公网侧负载分担节点将报文的外层公网IP地址以及该报文携带的用户信息结合考虑，选择转发路径。

如图2所示，为本发明实施例隧道负载分担的方法的流程图，具体包括:

步骤 S201，接收隧道的入口节点发送的报文，该报文的报文头中包含用户信息。

在本发明实施例中，隧道的入口节点根据用户的 IP 地址、MAC 地址和会话标识中的一种或几种，进行哈希获取该用户的用户信息，并将该用户信息封装在报文头的 Options 域中。

该 Options 域包括选项类型字段、选项长度字段和选项数据字段，其中，选项数据字段承载获取的用户信息。

但是本发明实施例在获取用户信息时并不局限于此，还可以对用户的其他标识信息进行哈希获取用户的用户信息。

步骤 S202，根据用户信息和报文头中的公网 IP 地址，进行选路，转发报文。

当需要对所述报文进行分片时，负载分担节点要将包含用户信息的 Options 域复制到分片后每个报文的报文头中。

上述隧道负载分担的方法，负载分担节点接收包含用户信息的报文，并根据报文头中的IP地址和用户信息进行选路，转发该报文。因此，在隧道数目有限或隧道地址分散时，负载分担节点也可以将流量均匀地分布到每条链路上。并且在对报文进行分片时，将包含用户信息的Options域复制到分片后每个报文的报文头中，从而避免了报文乱序问题的产生。

如图3所示，为本发明实施例隧道负载分担的方法的组网示意图，本发明实施例以IPv4网络为例进行说明，图3中，路由器A为隧道的入口节点、路由器B为公网侧的负载分担节点、路由器C为公网侧普通的转发节点和路由器D为隧道的出口节点。下面将详细介绍每个路由器进行的操作。

如图4所示，为本发明实施例隧道的入口节点的操作流程示意图，本发明实施例中，隧道的入口节点为路由器A。具体包括:

步骤S401，隧道的入口节点配置是否将用户信息携带到公网侧的指示信息。该指示信息指示隧道的入口节点是否将用户信息携带到公网侧，当该指示信息指示隧道的入口节点将用户信息携带到公网侧时，该指示信息还可以进一步指示用户信息的获取方式。具体可以为:

（1）根据用户的IP地址进行HASH获取用户信息；或者，

（2）根据用户的MAC地址进行HASH获取用户信息；或者，

（3）根据隧道建立时的会话标识（例如：L2TP隧道）进行HASH获取用户信息。

另外，在获取用户信息时，还可以将上述3种方式两两进行组合，或者同时使用上述3种方式获取用户信息，即本发明实施例可以根据用户的IP地址、MAC地址和会话标识中的一种或几种，进行HASH获取用户信息。

但是本发明实施例在获取用户信息时并不局限于上述3种方式，还可以对用户的其他标识信息进行HASH获取用户信息。

步骤S402，对用户报文进行入隧道封装时，获取用户信息。

在对用户报文进行入隧道封装时，当步骤S401中配置的指示信息指示隧道的入口节点将用户信息携带到公网侧时，隧道的入口节点根据该指示信息所指示的方式，获取用户信息。

步骤S403，在封装公网侧IPv4地址时，将步骤S402计算得到的用户信息封装在公网侧的IPv4报文头中。本发明实施例使用IPv4报文头中的选项（Options）域携带用户信息，包括选项（Options）域的IPv4报文头的结构如图5所示。

本发明实施例中的Options域的结构如图6所示，该Options域带有一个字节的选项类型（Option-type）字段，一个字节的选项长度（Option-length）字段和多个字节的选项数据字段。

其中，选项类型（Option-type）字段包含3个域:

（1）拷贝（copied）域：长度为1比特，在有分片的情形时，指示是否将Options域拷贝到分片报文中，例如：当拷贝域的值为1时，指示将Options域拷贝到分片报文中；当拷贝域的值为0时，指示不将Options域拷贝到分片报文中。

（2）类别（class）域：长度为2比特，当类别域的值为0时，表示控制；当类别域的值为2时，表示调试和度量；1和3为预留值，以备将来使用。

（3）选项号（number）域：长度为5比特，为每一种选项设置一种选项号用以识别这种选项。

本发明实施例中将拷贝域设置为1，将类别域设置为0，选项号定义为6。因此在本发明实施例中，选项类型字段的值为134，表示Options域中包含负载分担参数。

其中，选项长度字段的值是选项类型字段、选项长度字段和选项数据字段长度的总和。由于本发明实施例中携带的用户信息是为了在负载分担节点进行选路，而实际组网中链路的数目不会很多，同时为了保证IP头是32比特的整数倍，因此，本发明实施例将选项数据字段的长度设为8比特，该选项数据

字段携带哈希后的用户信息（HASH_INFORMATION）。这时，选项长度字段的值为4。

对于公网侧的负载分担节点路由器B，在接收到路由器A发送的IPv4报文之后，对该IPv4报文的报文头进行解析，查找路由，在发现需要进行负载分担时，路由器B判断IPv4报文头中是否带有Options域。当IPv4报文头中包含Options域时，路由器B进一步判断Options域的选项类型字段中是否包含负载分担参数，如果包含，则路由器B根据Options域的选项数据字段中哈希后的用户信息（HASH_INFORMATION），以及公网IPv4地址进行选路，转发上述报文，同时对Options域不作修改。

当需要对IPv4报文进行分片时，IPv4报文头中的Options域也会被复制到分片报文的IPv4报文头中。

对于公网侧普通的转发节点路由器C，支持对包含Options域的IPv4报文头的解析，对包含上述IPv4报文头的IPv4报文进行正常转发，不需要考虑IPv4报文头中的Options域，对Options域也不作任何修改，将包含上述IPv4报文头的IPv4报文透传到下一个节点。

当需要对IPv4报文进行分片时，IPv4报文头中的Options域也会被复制到分片报文的IPv4报文头中。

对于隧道的出口路由器D，支持对包含Options域的IPv4报文头的解析，正常解析包含上述IPv4报文头的IPv4报文，当该IPv4报文为出隧道报文时，路有器D对该IPv4报文进行解封装，剥去外层的IPv4报文头，然后对该IPv4报文进行出隧道处理。

如图 7 所示，为本发明实施例隧道负载分担的系统的结构图，包括:

隧道的入口节点 71，用于获取用户的用户信息，将用户信息封装在报文头中，并发送携带所述报文头的报文;

负载分担节点 72，用于接收隧道的入口节点 71 发送的报文，根据报文头中的用户信息和公网 IP 地址，进行选路，转发所述报文。

该负载分担的系统还可以包括：转发节点 73，用于透传负载分担节点 72 转发的报文;

隧道的出口节点 74，用于接收转发节点 73 透传的报文，解析该报文，当报文为出隧道报文时，对所述报文进行解封装，去掉该报文的报文头，并对所述报文进行出隧道处理。

上述隧道负载分担的系统，负载分担节点 72 接收隧道的入口节点 71 发送的包含用户信息的报文，并根据报文头中的 IP 地址和用户信息进行选路，转发该报文。因此，在隧道数目有限或隧道地址分散时，负载分担节点 72 也可以将流量均匀地分布到每条链路上。

如图 8 所示，为本发明实施例隧道的入口节点的结构图，包括:

获取模块 711，用于获取用户的用户信息;

封装模块 712，用于将获取模块 711 获取的用户信息封装在报文头中;

发送模块 713，用于发送携带封装模块 712 封装的报文头的报文。

其中，获取模块 711 可以包括: 哈希子模块 7111，用于根据用户的 IP 地址、MAC 地址和会话标识中的一种或几种，进行哈希获取所述用户的用户信息。

上述隧道的入口节点，获取模块 711 获取用户的用户信息，封装模块 712 将获取模块 711 获取的用户信息封装在报文头中，由发送模块 713 发送携带上述报文头的报文，从而实现了将用户信息携带到公网侧，使公网侧的负载分担节点 72 在进行负载分担时，可以根据用户信息和公网 IP 地址进行选路。

如图 9 所示，为本发明实施例负载分担节点的结构图，包括:

接收模块 721，用于接收隧道的入口节点 71 发送的报文，该报文的报文头中包含用户信息;

报文转发模块 722，用于根据接收模块 721 接收的报文头中的用户信息和公网 IP 地址，进行选路，转发所述报文。

其中，负载分担节点 72 还可以包括: 复制模块 723，用于在对接收模块 721 接收的报文进行分片时，将接收模块 721 接收的报文头中的 Options 域复制到分片后每个报文的报文头中，该 Options 域携带所述用户信息。

上述负载分担节点，接收模块 721 接收包含用户信息的报文，报文转发模块 722 根据报文头中的 IP 地址和用户信息进行选路，转发该报文。因此，

在隧道数目有限或隧道地址分散时，负载分担节点 72 也可以将流量均匀地分布到每条链路上。并且在对报文进行分片时，复制模块 723 将包含用户信息的 Options 域复制到分片后每个报文的报文头中，从而避免了报文乱序问题的产生。

本发明实施例解决了IPv4公网侧由于隧道数目有限或隧道地址分散，导致负载分担节点无法将流量均匀分布到每条链路上的问题。本发明实施例将用户信息带到公网侧，使之成为进行负载分担的依据之一，并不改变隧道对用户报文保密的属性。同时本发明实施例不限定隧道的具体模型，针对承载于IPv4网络的所有隧道。本发明实施例中，负载分担节点可以根据用户或隧道模型的需要，灵活地使用用户的IP地址、MAC地址、会话标识等信息进行选路，实现IPv4了公网侧的灵活负载分担。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到本发明可以通过硬件实现，也可以可借助软件加必要的通用硬件平台的方式来实现基于这样的理解，本发明的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）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）执行本发明各个实施例所述的方法。

本领域技术人员可以理解附图只是一个优选实施例的示意图，附图中的模块或流程并不一定是实施本发明所必须的。

本领域技术人员可以理解实施例中的装置中的模块可以按照实施例描述进行分布于实施例的装置中，也可以进行相应变化位于不同于本实施例的一个或多个装置中。上述实施例的模块可以合并为一个模块，也可以进一步拆分成多个子模块。

上述本发明实施例序号仅仅为了描述，不代表实施例的优劣。

以上公开的仅为本发明的几个具体实施例，但是，本发明并非局限于此，任何本领域的技术人员能思之的变化都应落入本发明的保护范围。

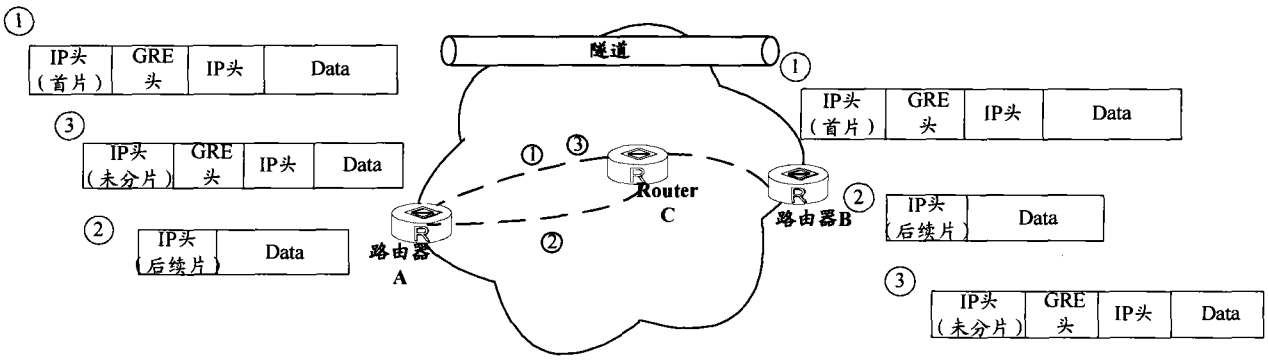

①
IP头（首片）
GRE头
IP头
Data
③
IP头（未分片）
GRE头
IP头
Data
②
IP头后续片
Data
隧道
①
③
②
Router C
路由器A
路由器B
①
IP头（首片）
GRE头
IP头
Data
②
IP头后续片
Data
③
IP头（未分片）
GRE头
IP头
Data

图 1

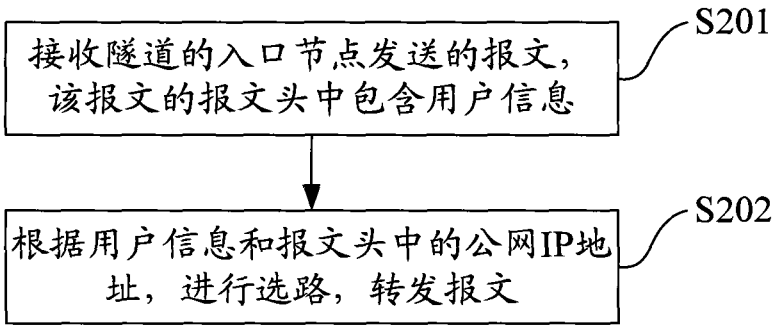

接收隧道的入口节点发送的报文，该报文的报文头中包含用户信息
S201
根据用户信息和报文头中的公网IP地址，进行选路，转发报文
S202

图 2

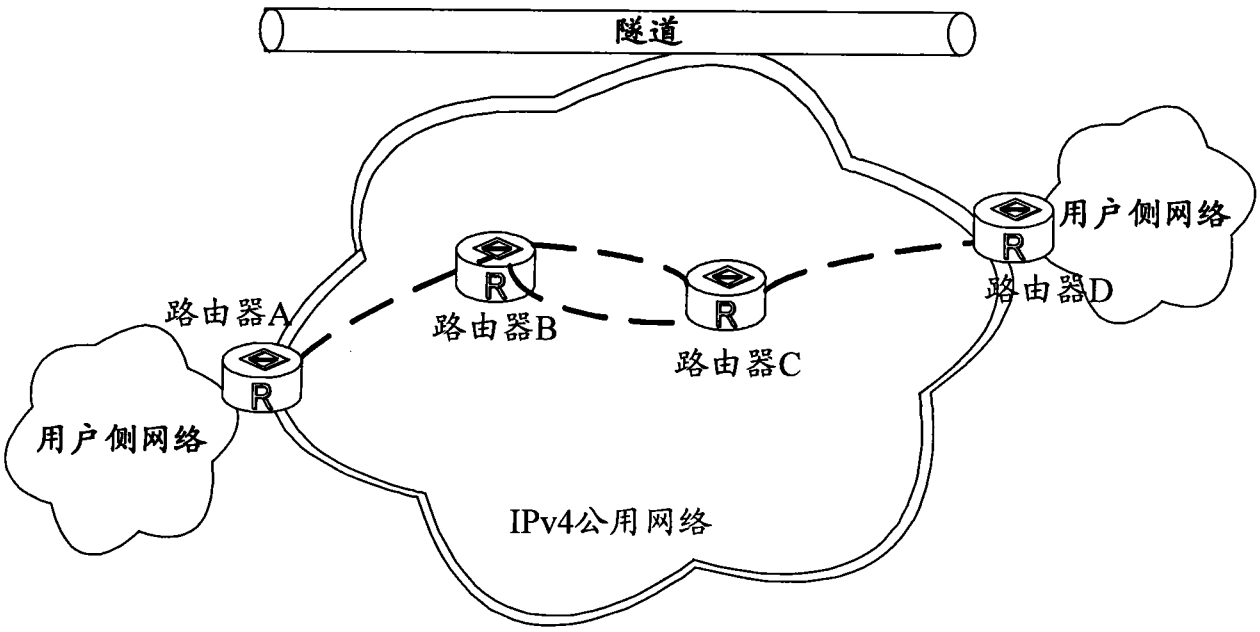

隧道
用户侧网络
路由器A
路由器B
路由器C
路由器D
用户侧网络
IPv4公用网络

图 3

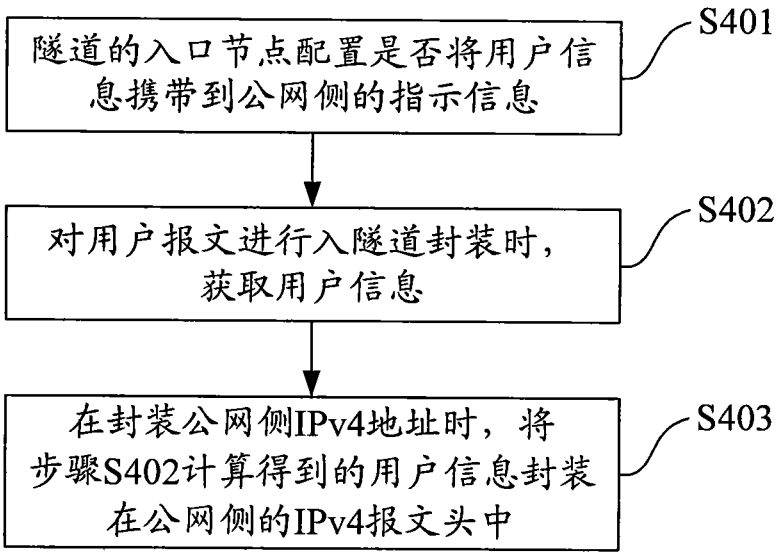

隧道的入口节点配置是否将用户信息携带到公网侧的指示信息
S401
对用户报文进行入隧道封装时，获取用户信息
S402
在封装公网侧IPv4地址时，将步骤S402计算得到的用户信息封装在公网侧的IPv4报文头中
S403

图 4

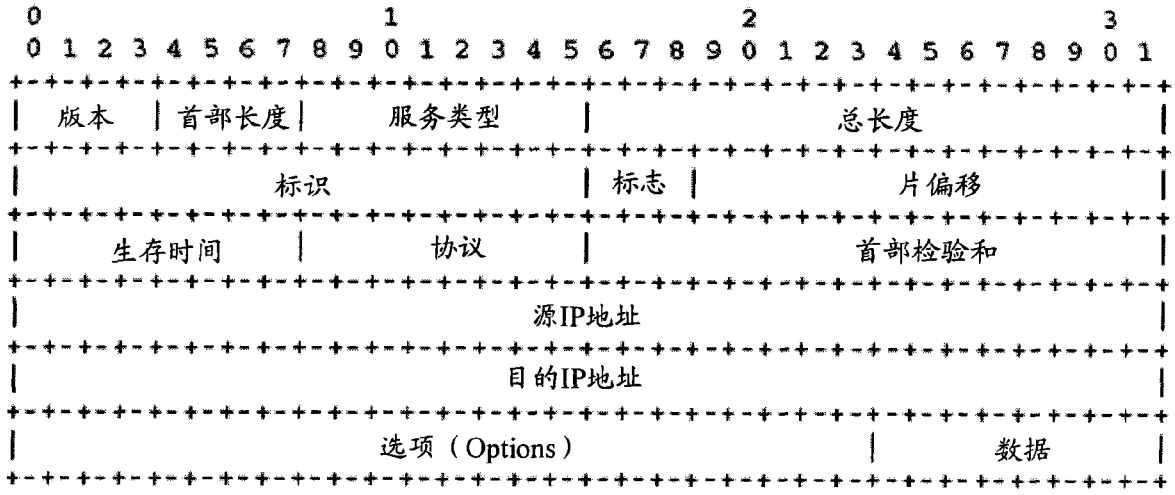
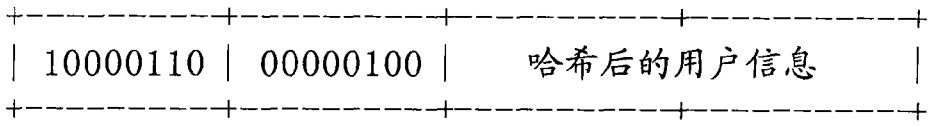


图 5



选项类型=134 选项长度＝4

图 6

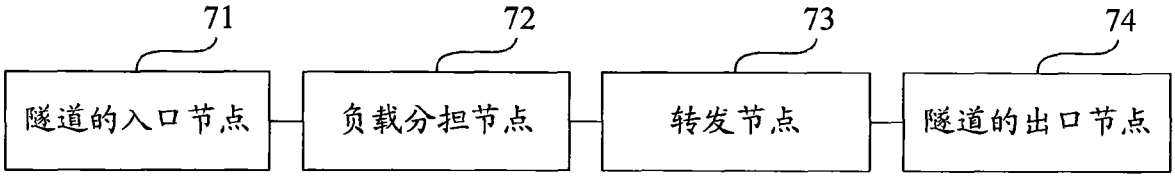


图 7

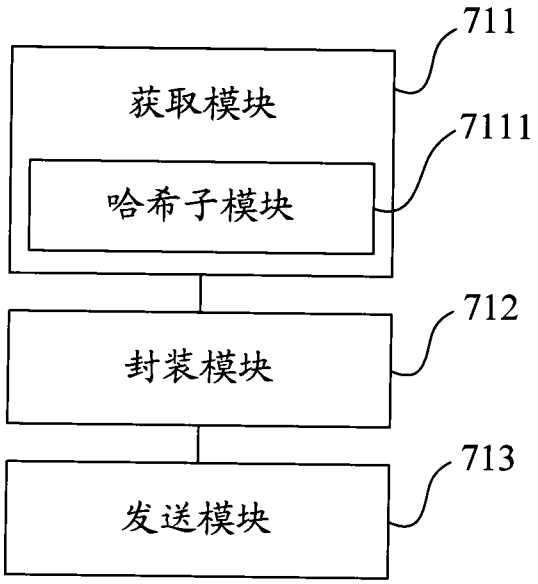

711
获取模块
7111
哈希子模块
712
封装模块
713
发送模块

图 8

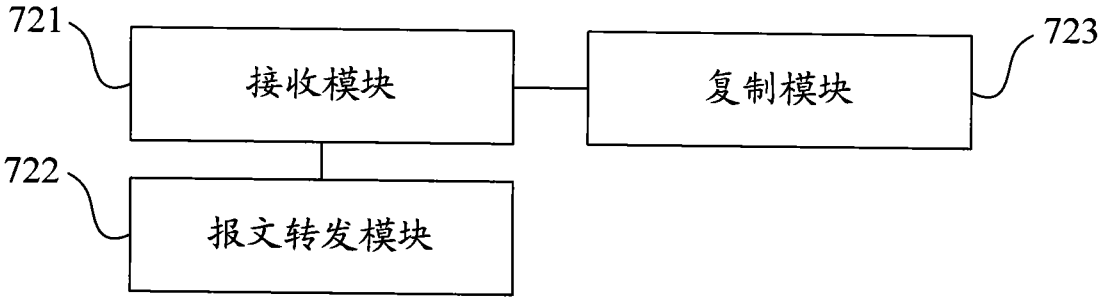

721
接收模块
723
复制模块
722
报文转发模块

图 9