[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04L 12/28 (2006.01)
H04L 29/06 (2006.01)
H04L 12/24 (2006.01)

[12] 发明专利说明书

专利号 ZL 03110053.8

[45] 授权公告日 2007 年 2 月 14 日

[11] 授权公告号 CN 1300986C

[22] 申请日 2003.4.14 [21] 申请号 03110053.8

[73] 专利权人 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 龚 华 熊 鹰

[56] 参考文献
CN 1392701A 2003.1.22
CN 1400535A 2003.3.5
JP 2002-281104A 2002.9.27
WO 03/015330A2 2003.2.20
WO 01/60025A2 2001.8.16
EP 1175042A2 2002.1.23
US 6327626B1 2001.12.4
审查员 鲁艳萍

[74] 专利代理机构 北京凯特来知识产权代理有限公司
代理人 郑立明

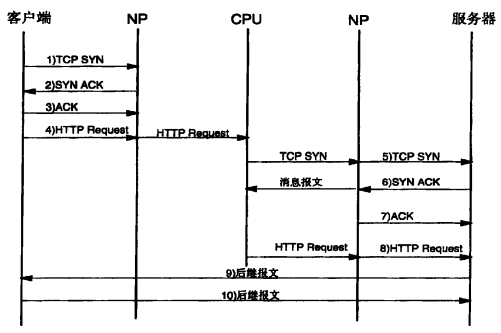
权利要求书 3 页 说明书 10 页 附图 4 页

[54] 发明名称

实现快速五七层交换的方法

[57] 摘要

一种实现快速五七层交换的方法包括：发送 TCP SYN；构造 SYNACK 报文；发送 ACK 报文；发送一个带有应用层信息的内容请求报文；根据报文状态及报文种类，将报文通过总线上送给 CPU；CPU 收到所述上送的内容请求报文后，提取应用层信息并根据配置的内容规则进行内容匹配，选择合适的服务器组，构造 TCP SYN 报文下发；将 TCP SYN 报文发送给真实服务器；发送 SYN ACK 报文，构造 ACK 报文，构造消息报文通过总线上送给 CPU；将缓存的 HTTP 请求报文下发，将 HTTP 请求报文转发给服务器；直接转发后继报文。本发明有效地减少 NP 与 CPU 交互的报文，减轻了 CPU 的负担。



1、一种实现快速五七层交换的方法，包括步骤：

客户端发送TCP SYN;

网络处理器NP收到该TCP SYN报文之后，构造SYN ACK报文，对客户端进行响应，NP为客户端侧后继报文建立一条状态为TCP哄骗的流Cache表项;

客户端收到来自NP的SYN ACK报文之后，向NP发送ACK报文;

客户端发送一个带有应用层信息的内容请求报文;

NP根据报文状态及报文种类，将报文通过总线上送给CPU;

CPU收到所述上送的内容请求报文后，提取应用层信息并根据配置的内容规则进行内容匹配，选择合适的服务器组，构造TCP SYN报文下发给NP;

NP将TCP SYN报文发送给真实服务器;

服务器收到所述TCP SYN之后，响应客户端的请求发送SYN ACK报文，NP根据报文状态生成ACK报文响应服务器；更新两侧流Cache表项；构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU改造HTTP请求报文，并下发给NP;

NP将HTTP请求报文转发给服务器;

NP直接转发后继报文。

2、如权利要求1所述的方法，其中，所述客户端收到来自NP的SYN ACK报文之后，向NP发送ACK报文的步骤还包括步骤：所述ACK报文到达NP后命中流Cache，NP根据流Cache的状态以及报文的种类做出丢弃决定。

3、如权利要求2所述的方法，其中所述客户端发送的一个带有应用层信息的内容请求报文到达NP之后同样会命中流Cache；NP根据流Cache的状态以及报文种类做出上送CPU的决定，将报文通过总线上送给CPU。

4、如权利要求2所述的方法，其中所述CPU收到所述上送的内容请求报文后，提取应用层信息并根据配置的内容规则进行内容匹配，选择合适的服务器组，构造TCP SYN报文下发给NP的步骤包括步骤：CPU收到所述上送的内容请求报文后，建一个TCP控制块记录该报文的基本信息，并将该报文缓存。

5、如权利要求2所述的方法，其中所述NP将TCP SYN报文发送给真实服务器的步骤包括步骤：

进行负载均衡调度；

选择一台真实服务器；

用真实服务器的IP地址替换CPU构造的TCP SYN报文中的目的IP地址；

计算IP头校验和与TCP校验和；

接着建立一条状态为TCP哄骗的服务器侧流Cache；

记录TCP控制块的序号。

6、如权利要求5所述的方法，其中所述进行负载均衡调度包括在服务器组中按照加权轮转、加权最小连接数、哈希负载均衡之一或其组合进行负载均衡调度。

7、如权利要求5所述的方法，其中所述服务器收到所述TCP SYN之后，响应客户端的请求发送的SYN ACK报文到达NP后会命中流Cache。

8. 如权利要求5所述的方法，包括：所述NP根据流Cache的状态生成ACK报文响应服务器；更新两侧流Cache，其中流Cache状态更新为直接转发；构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU

改造先前缓存的HTTP请求报文，并下发给NP；且由NP直接转发两侧的后继报文命中流Cache。

9、如权利要求5所述的方法，其中所述服务器收到所述TCP SYN之后，响应客户端的请求发送的SYN ACK报文到达NP后会命中流Cache，NP根据流Cache的状态生成ACK报文响应服务器；更新两侧流Cache，其中流Cache状态更新为上送CPU；构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU改造先前缓存的HTTP请求报文，并下发给NP；且由NP直接转发两侧的后继报文命中流Cache。

10、如权利要求9所述的方法，还包括步骤：

服务器收到加密套接字协议层SSL内容请求报文之后，发送带有所述SSL信息的响应报文，所述报文到达NP之后命中所述流Cache，NP根据所述流Cache的状态将报文上送给CPU;

CPU提取SSL信息，判断其合法性，建立SSL信息与真实服务器一一对应的对应关系的表进行维护;

改造所述SSL报文，重新计算校验和，

下发所述报文给NP，由NP将报文转发给客户端;

CPU会下发一个更新流Cache的消息报文，以将两侧流Cache的状态更新为直接转发。

实现快速五七层交换的方法

技术领域

本发明涉及IP（Internet Protocol网际协议）通信，具体涉及实现快速五七层交换的方法。

背景技术

为了叙述的方便，本说明书中的下列短语的定义如下：

NP　　Network Processor，网络处理器

五七层交换　　多层交换机通过感知报文的应用层信息，根据内容完成报文的交换过程

CPU　　Central Processing Unit 中央处理器

IP　　Internet Protocol 网际协议

TCP　　Transfer Control Protocol 传输控制协议

TCP SYN　　SYN是 同步序号标志，TCP首部中的一个标志位。当新建一个TCP连接的时候，请求端（通常称为客户端）需要首先发送一个置了SYN标志的TCP报文。

SYN ACK　　ACK是确认标志，TCP首部中的一个标志位。SYN ACK在本文中表示同时置上了这两个标志位的TCP报文，是服务器响应TCP SYN而发送的确认报文。

ACK　　表示仅置了ACK标志的TCP报文，是客户端响应SYN ACK而发送的确认报文。此报文发送后，一个TCP连接就完成了。这个过程也称为三次握手。

HTTP Request　　内容请求报文，本文泛指在完成TCP三次握手之后，客户端紧接着发送的含有应用层信息的TCP报文。

HTTP　　Hypertext Transfer Protocol，万维网服务程序所用的协议

Cookie 一种网络服务器传递给浏览器的信息，用于实现粘性连接

SYN FLOOD 一种拒绝服务的攻击手段，通过发送大量没有后继报文的TCP SYN报文，来达到消耗目标服务器或者交换机的资源，使之不能提供正常服务。

SSL Security Socket Layer 加密套接字协议层

真实服务器 能提供具体服务的服务器

服务器组 若干真实服务器的集合

五七层交换是利用应用层信息来识别应用数据流会话，根据配置的内容交换规则来决定报文的转发。为了截获客户端数据包的应用层信息，转发设备采用TCP哄骗的技术来分别完成与客户端和服务器的TCP三次握手过程，所以完成一次内容交换（真实服务器收到含有内容请求的报文，如图1，转发设备就要处理8个报文。

转发设备的不同以及设备内部处理的不同就构成了现有五七层交换技术的不同。

值得一提的是，不同的技术对SYN FLOOD攻击的抵抗能力也是截然不同的。所谓的SYN FLOOD攻击就是：恶意攻击者利用某种手段构造大量的目的IP地址为目标服务器的TCP SYN报文（没有后继报文），以此来达到消耗目标服务器的CPU资源，使目标服务器不能提供正常服务的目的。这种攻击对中间转发设备同样适用。

现有技术一采用软交换技术，全部处理都在CPU上完成。即虚拟服务器方案。图2描述现有技术一采用虚拟服务器五七层交换TCP完成一次内容交换转发的信号流程图。所有的TCP哄骗与内容匹配工作均由高性能CPU完成。其优点是实现简单，成本较低。但由于该技术没有用NP，所以转发性能差，只能带少量的服务器进行负载均衡。对SYN FLOOD攻击的抵抗能力很差。

现有技术二中采用网络处理器，通过NP与CPU的配合实现五七层交换，但把TCP哄骗和内容匹配等大部分工作都交给CPU做，NP负责将报文

上送给CPU并负责报文的转发。图3是现有技术二的系统结构图。其中的NP就是网络处理器，其分布式结构设计与多线程并发处理可以实现高性能的报文转发。NP与CPU通过总线进行通信。图4描述现有技术二采用多层交换机五七层交换TCP完成一次内容交换转发的信号流程图。其信号处理流程如下:

1) NP收到客户端的TCP SYN报文，将该报文上送给CPU;

2)CPU构造TCP SYN ACK报文下发给NP，由NP转发给客户端，同时NP为客户端添加一条流Cache表项（该表项记录了该TCP流的基本信息以及处理信息）;

3)NP收到客户端的TCP ACK报文，该报文命中流Cache，获取相关信息后，将报文上送给CPU; CPU丢弃该报文，并进行状态迁移; 至此完成了客户端的TCP哄骗。

4)NP收到客户端的HTTP请求报文，该报文命中流Cache，获取相关信息后，将报文上送给CPU; CPU提取报文的应用层信息，根据配置的内容规则选择适当地内容服务器组; 然后通过一定的负载均衡调度策略在内容服务器组中选择合适的真实服务器; 缓存该报文，并构造去往该真实服务器的TCP SYN报文，将TCP SYN报文下发给NP;

5)NP将TCP SYN报文转发给该真实服务器; 同时为服务器添加一条流Cache表项;

6)NP收到服务器的TCP SYN ACK报文，该报文命中流Cache，获取相关信息后，NP将该报文上送CPU;

7)CPU收到该报文后，构造TCP ACK报文下发给NP，由NP将ACK报文转发给服务器; 至此服务器端的TCP哄骗完成。

8)CPU将缓存的HTTP Request报文修改后，下发给NP，由NP负责转发给服务器; 同时下发控制帧更新两侧的流Cache表项。

至此，整个HTTP内容交换的主要工作就完成了，该TCP流的后继报文会命中流Cache并直接由NP进行转发。

由于采用了高性能的网络处理器，其性能已经有了质的飞跃。但从系统结构原理图中可以看出，NP与CPU之间的通信是通过总线来完成的，所以不可避免的成为系统的瓶颈。而且在该方案中，完成一次TCP流的五七层交换NP与CPU交互的报文至少要8个，势必极大的影响性能。再加上CPU还要完成TCP哄骗，性能就更差了。从安全性方面考虑，一旦遭受SYN FLOOD 攻击，CPU要为每个连接保存状态而且不能正常释放，所以CPU的资源会很快被耗尽，以至不能提供正常的服务。

发明内容

为了解决现有技术的不足，本发明采用TCP哄骗的大部分工作以及负载均衡调度可以交给NP来完成。这样就能有效的减少NP与CPU交互的报文，而且减轻了CPU的负担。

本发明提供了一种实现快速五七层交换的方法，包括步骤:

客户端发送TCP SYN;

NP收到该TCP SYN报文之后，构造SYN ACK报文，对客户端进行响应，NP为客户端侧后继报文建立一条状态为TCP哄骗的流Cache表项;

客户端收到来自NP的SYN ACK报文之后，向NP发送ACK报文;

客户端发送一个带有应用层信息的内容请求报文;

NP根据报文状态及报文种类，将报文通过总线上送给CPU;

CPU收到所述上送的内容请求报文后，提取应用层信息并根据配置的内容规则进行内容匹配，选择合适的服务器组，构造TCP SYN报文下发给NP;

NP将TCP SYN报文发送给真实服务器;

服务器收到所述TCP SYN之后，响应客户端的请求发送SYN ACK报文，NP根据报文状态生成ACK报文响应服务器；和/或更新两侧报文；和/或构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU改造HTTP请求报文，并下发给NP;

NP将HTTP请求报文转发给服务器;

NP直接转发后继报文。

可选地，所述客户端收到来自NP的SYN ACK报文之后，向NP发送ACK报文的步骤还包括步骤：所述ACK报文到达NP后命中流Cache，NP根据流Cache的状态以及报文的种类做出丢弃决定。

优选地，所述客户端发送的一个带有应用层信息的内容请求报文到达NP之后同样会命中流Cache；NP根据流Cache的状态以及报文种类做出上送CPU的决定，将报文通过总线上送给CPU。

可选地，所述CPU收到所述上送的内容请求报文后，提取应用层信息并根据配置的内容规则进行内容匹配，选择合适的服务器组，构造TCP SYN报文下发给NP的步骤包括步骤：CPU收到所述上送的内容请求报文后，建一个TCP控制块记录该报文的基本信息，并将该报文缓存。

优选地，所述NP将TCP SYN报文发送给真实服务器的步骤包括步骤：

进行负载均衡调度；

选择一台真实服务器；

用真实服务器的IP地址替换CPU构造的TCP SYN报文中的目的IP地址；

计算IP头校验和与TCP校验和；

接着建立一条状态为TCP哄骗的服务器侧流Cache；

记录TCP控制块的序号。

可选地，所述进行负载均衡调度包括在服务器组中按照加权轮转、加权最小连接数、哈希负载均衡。

优选地，所述服务器收到所述TCP SYN之后，响应客户端的请求发送的SYN ACK报文到达NP后会命中流Cache，NP根据流Cache的状态生成ACK报文响应服务器；更新两侧流Cache，其中流Cache状态更新为直接转发；构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU改造先前缓存的HTTP请求报文，并下发给NP；且其中所述由NP直接转发两侧的后继报文命中流Cache。

可选地，所述服务器收到所述TCP SYN之后，响应客户端的请求发送的SYN ACK报文到达NP后会命中流Cache，NP根据流Cache的状态生成ACK报文响应服务器；更新两侧流Cache，其中流Cache状态更新为上送CPU；构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU改造先前缓存的HTTP请求报文，并下发给NP；且其中所述由NP直接转发两侧的后继报文命中流Cache。

优选地，该方法还包括步骤：

服务器收到SSL内容请求报文之后，发送带有所述SSL信息的响应报文，所述报文到达NP之后命中所述流Cache，NP根据所述流Cache的状态将报文上送给CPU；

CPU提取SSL信息，判断其合法性，建立维护SSL信息与真实服务器的对应关系的表（一一对应）；

改造所述SSL报文，重新计算校验和，

下发所述报文给NP，由NP将报文转发给客户端；

CPU会下发一个更新流Cache的消息报文，以将两侧流Cache的状态更新为直接转发。

利用本发明，TCP哄骗的大部分工作以及负载均衡调度可以交给NP来完成。这样就能有效的减少NP与CPU交互的报文，而且减轻了CPU的负担。

附图说明

图1描述TCP完成一次内容交换转发的信号流程图；

图2描述现有技术一采用虚拟服务器五七层交换转发的信号流程图；

图3是现有技术二的系统结构图；

图4描述现有技术二采用多层交换机五七层内容交换转发的信号流程图；

图5描述本发明的采用多层交换机五七层内容交换转发的信号流程图；

图6描述本发明的采用多层交换机五七层交换实现比较复杂的SSL粘性连接的信号流程图；

具体实施方式

本发明是对现有技术二的改进，在本发明中，采用NP来处理一些现有技术二由CPU处理的工作，TCP哄骗的大部分工作以及负载均衡调度均交给NP来完成。这样就能有效的减少NP与CPU交互的报文，而且减轻了CPU的负担。

在本发明中，整个五七层交换过程由流Cache（高速缓存）表进行状态控制，一个TCP流分别对应客户端侧Cache和服务器侧Cache两条流Cache表项，每条表项分为三个状态：TCP哄骗、上送CPU、直接转发。

图5描述本发明的采用多层交换机五七层内容交换转发的信号流程图。在本发明中，五七层交换过程的具体步骤如下：

在步骤1，客户端首先发送TCP SYN，NP收到该TCP SYN报文之后，不向CPU转发，由NP直接构造SYN ACK报文，然后由NP进行转发响应客户端，同时为客户端侧后继报文建立一条流Cache表项，此时的状态为TCP哄骗。

然后，在步骤2，客户端收到来自NP的SYN ACK报文之后，马上向NP发送ACK报文，该报文到达NP之后会命中流Cache，然后，NP根据流Cache的状态以及报文的种类做出丢弃决定。

在步骤3，客户端在发送完ACK报文之后，紧接着会发送一个带有应用层信息的内容请求报文，该报文到达NP之后同样会命中流Cache，NP根据流Cache的状态以及报文种类做出上送CPU的决定，将报文通过总线上送给CPU。

在步骤4，CPU收到该内容请求报文之后，新建一个TCP控制块记录该报文的基本信息，并将该报文缓存；然后提取应用层信息并根据配置的内容规则进行内容匹配，选择合适的服务器组，接着构造TCP SYN报文下发给NP。

在步骤5，NP首先要进行负载均衡调度，在服务器组中按照加权轮转、加权最小连接数、哈希等之一或其组合负载均衡策略选择一台真实服务器，然后用真实服务器的IP地址替换CPU构造的TCP SYN报文中的目的IP地址，并计算IP头校验和与TCP校验和；接着建立一条服务器侧流Cache，其状态为TCP哄骗，并记录TCP控制块的序号；最后将TCP SYN报文发送给真实服务器。

在步骤6，服务器收到TCP SYN之后，会响应客户端的请求并发送SYN ACK报文，该报文到达NP后会命中流Cache，NP根据流Cache的状态做以下三件事：a、生成ACK报文响应服务器；b、更新两侧流Cache，其中流Cache状态更新为直接转发；c、构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU改造先前缓存的HTTP请求报文，并下发给NP。

在步骤7，NP将HTTP请求报文转发给服务器。

在步骤8，两侧的后继报文将命中流Cache，并由NP直接转发。

图6描述本发明的采用多层交换机五七层交换实现比较复杂的SSL（加密套接字协议层）粘性连接的信号流程图；

在步骤1，客户端首先发送TCP SYN，NP收到该TCP SYN报文之后，不向CPU转发，由NP直接构造SYN ACK报文，然后由NP进行转发响应客户端，同时为客户端侧后继报文建立一条流Cache表项，此时的状态为TCP哄骗。

然后，在步骤2，客户端收到来自NP的SYN ACK报文之后，马上向NP发送ACK报文，该报文到达NP之后会命中流Cache，然后，NP根据流Cache的状态以及报文的种类做出丢弃决定。

在步骤3，客户端在发送完ACK报文之后，紧接着会发送一个带有应用层信息的内容请求报文，该报文到达NP之后同样会命中流Cache，NP根据流Cache的状态以及报文种类做出上送CPU的决定，将报文通过总线上送给CPU。

在步骤4，CPU收到该内容请求报文之后，新建一个TCP控制块记录该报文的基本信息，并将该报文缓存；然后提取应用层信息并根据配置的内容规则进行内容匹配，选择合适的服务器组，接着构造TCP SYN报文下发给NP。

在步骤5，NP首先要进行负载均衡调度，在服务器组中按照加权轮转、加权最小连接数、哈希等负载均衡策略选择一台真实服务器，然后用真实服务器的IP地址替换CPU构造的TCP SYN报文中的目的IP地址，并计算IP头校验和与TCP校验和；接着建立一条服务器侧流Cache，其状态为TCP哄骗，并记录TCP控制块的序号；最后将TCP SYN报文发送给真实服务器。

在步骤6，服务器收到TCP SYN之后，会响应客户端的请求并发送SYN ACK报文，该报文到达NP后会命中流Cache，NP根据流Cache的状态做以下三件事：a、生成ACK报文响应服务器；b、更新两侧流Cache，其中流Cache状态更新为上送CPU；c、构造消息报文，将服务器的IP地址以及序列号上送CPU，通知CPU改造先前缓存的HTTP请求报文，并下发给NP。

在步骤7，NP将HTTP请求报文转发给服务器。

在步骤8，服务器收到SSL内容请求报文之后，会发送带有SSL信息的响应报文，该报文到达NP之后命中流Cache，NP根据流Cache的状态将报文原封不动的上送给CPU；CPU提取SSL信息，并判断该信息的合法性，然后建立一张表来维护SSL信息与真实服务器的对应关系（一一对应）；接着改造SSL报文，重新计算校验和，将报文下发给NP，由NP将报文转发给客户端。同时CPU会下发一个更新流Cache的消息报文，将两侧流Cache的状态更新为直接转发。

在步骤9，两侧的后续报文均会命中流Cache，并由NP直接进行转发。

以上的处理流程是针对客户端第一次进行SSL访问的。当客户端保存了服务器的SSL信息之后，再次发起SSL连接，其处理流程与上面的处理流程基本相同。唯一的不同在于：CPU收到客户端的SSL内容请求报文

之后，可以提取客户端的SSL信息，然后通过查表就能得到上一次连接的真实服务器，将此信息通知NP，NP就不用再做负载均衡调度了。报文会送往客户端第一次建立连接的那台服务器。

虽然通过实施例描绘了本发明，本领域普通技术人员知道，本发明有许多变形和变化而不脱离本发明的精神，希望所附的权利要求包括这些变形和变化。

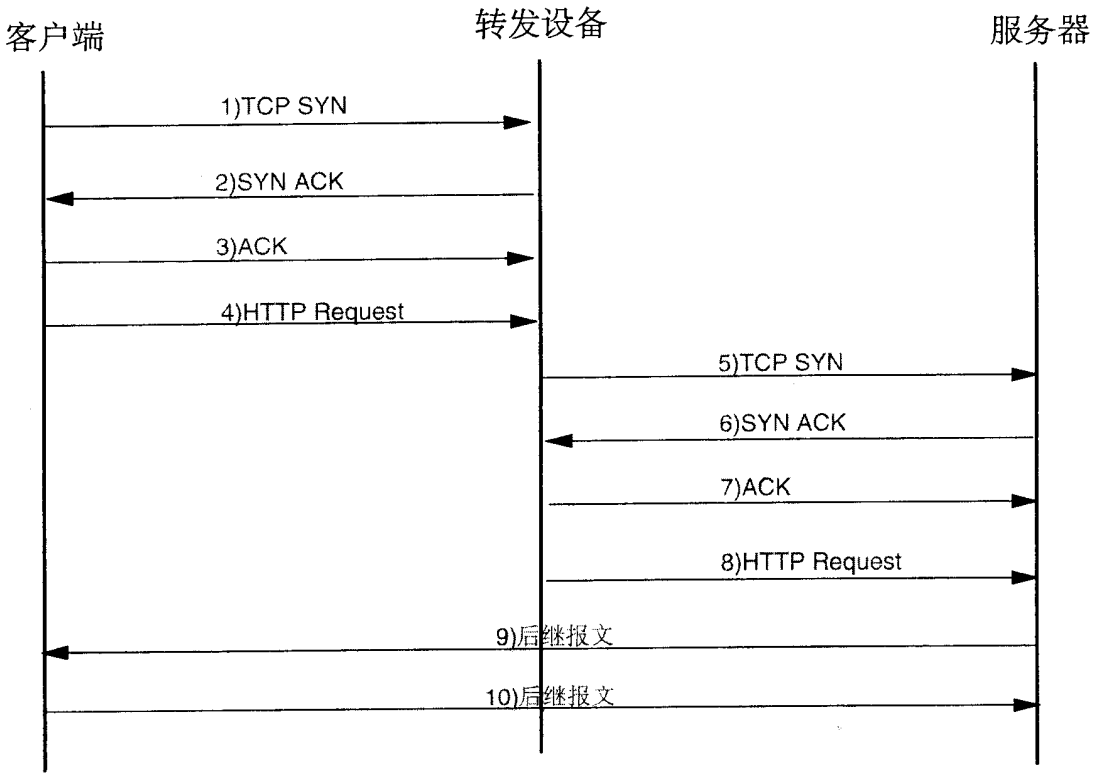

客户端
转发设备
服务器
1)TCP SYN
2)SYN ACK
3)ACK
4)HTTP Request
5)TCP SYN
6)SYN ACK
7)ACK
8)HTTP Request
9)后继报文
10)后继报文

图1

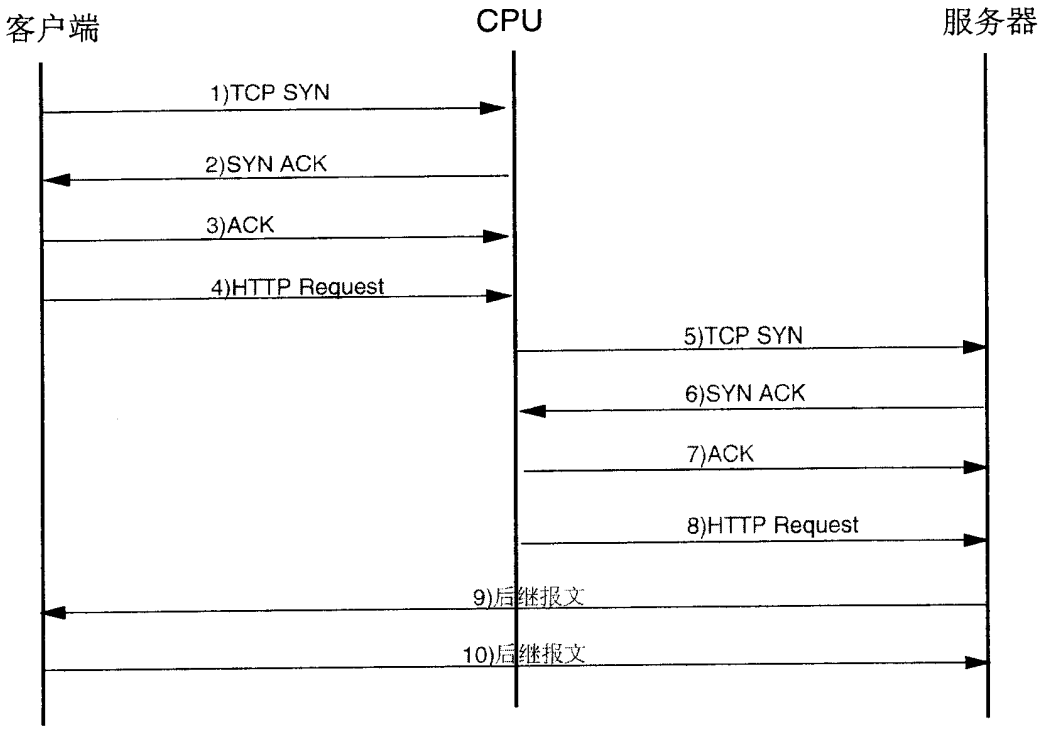

客户端
CPU
服务器
1)TCP SYN
2)SYN ACK
3)ACK
4)HTTP Request
5)TCP SYN
6)SYN ACK
7)ACK
8)HTTP Request
9)后继报文
10)后继报文

图2

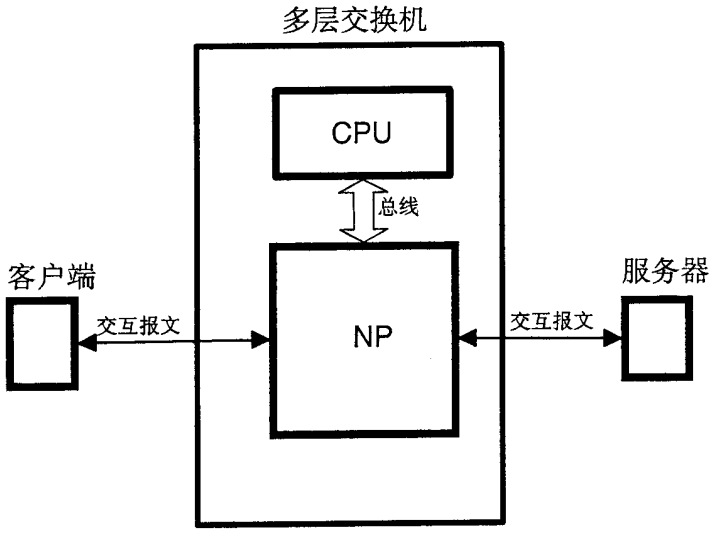

多层交换机
CPU
总线
客户端
交互报文
NP
交互报文
服务器

图3

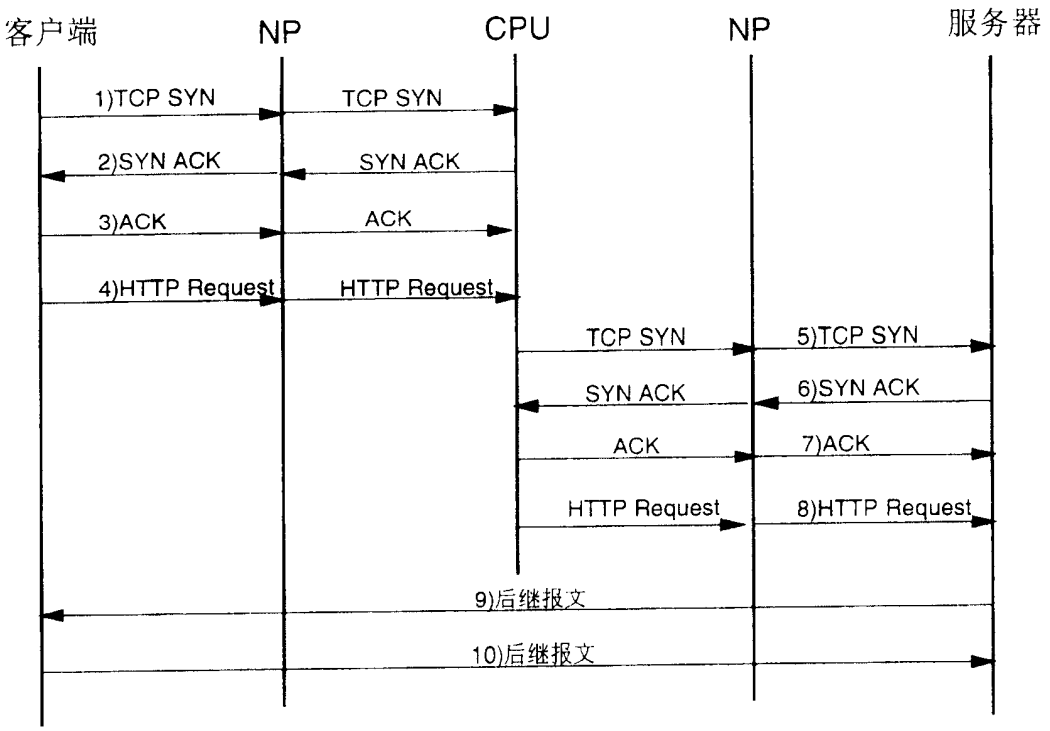

客户端
NP
CPU
NP
服务器
1)TCP SYN
TCP SYN
2)SYN ACK
SYN ACK
3)ACK
ACK
4)HTTP Request
HTTP Request
TCP SYN
5)TCP SYN
SYN ACK
6)SYN ACK
ACK
7)ACK
HTTP Request
8)HTTP Request
9)后继报文
10)后继报文

图4

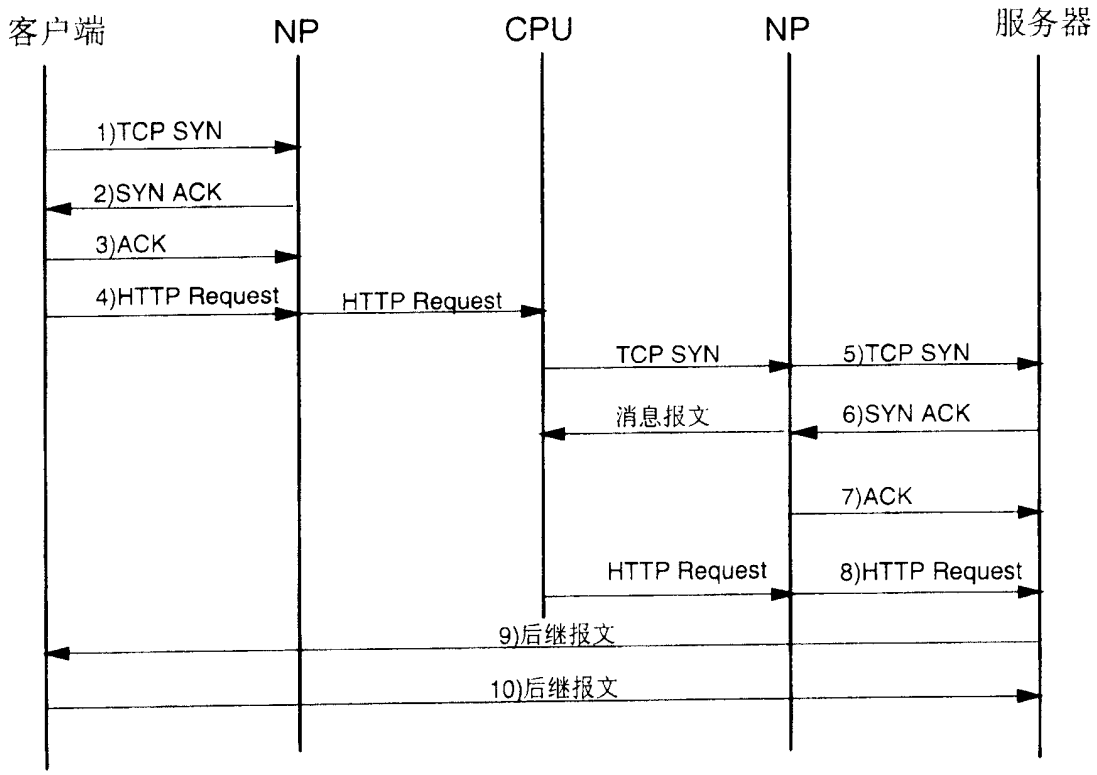

客户端
NP
CPU
NP
服务器
1)TCP SYN
2)SYN ACK
3)ACK
4)HTTP Request
HTTP Request
TCP SYN
5)TCP SYN
消息报文
6)SYN ACK
7)ACK
HTTP Request
8)HTTP Request
9)后继报文
10)后继报文

图5

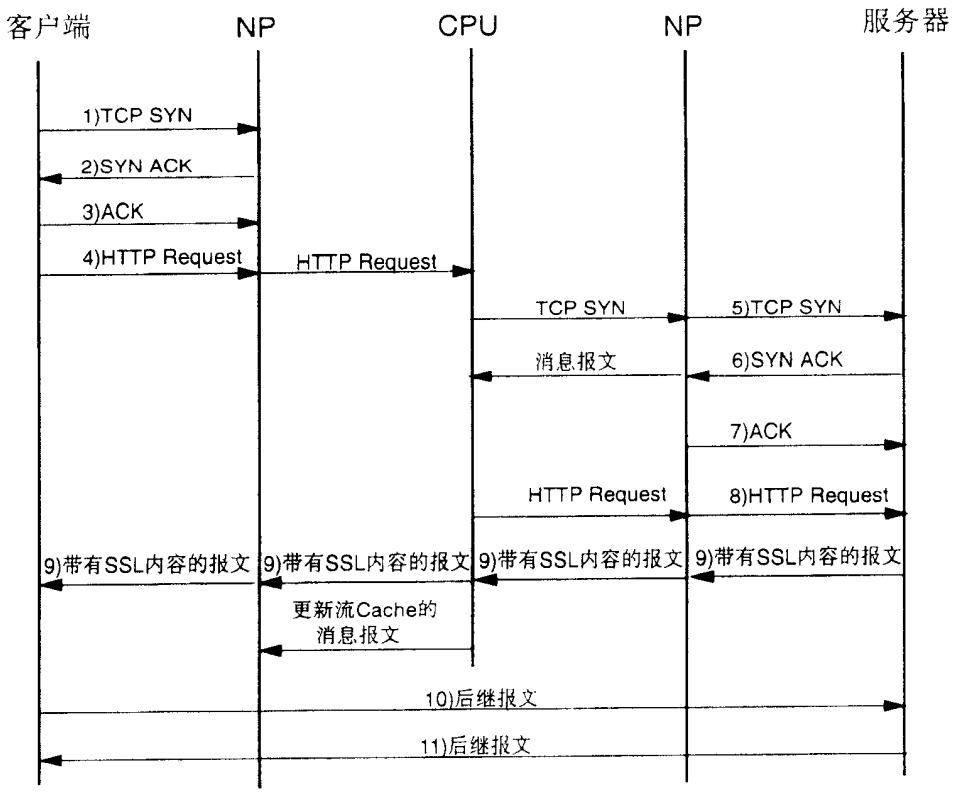

客户端
NP
CPU
NP
服务器
1)TCP SYN
2)SYN ACK
3)ACK
4)HTTP Request
HTTP Request
TCP SYN
5)TCP SYN
消息报文
6)SYN ACK
7)ACK
HTTP Request
8)HTTP Request
9)带有SSL内容的报文
9)带有SSL内容的报文
9)带有SSL内容的报文
9)带有SSL内容的报文
更新流Cache的
消息报文
10)后继报文
11)后继报文

图6