



(12) 发明专利申请

(10) 申请公布号 CN 102457440 A

(43) 申请公布日 2012. 05. 16

(21) 申请号 201210011400. 3

(22) 申请日 2012. 01. 13

(71) 申请人 华为软件技术有限公司  
地址 210012 江苏省南京市雨花台区安德门  
94 号

(72) 发明人 李小强

(51) Int. Cl.  
H04L 12/56 (2006. 01)  
H04L 29/12 (2006. 01)

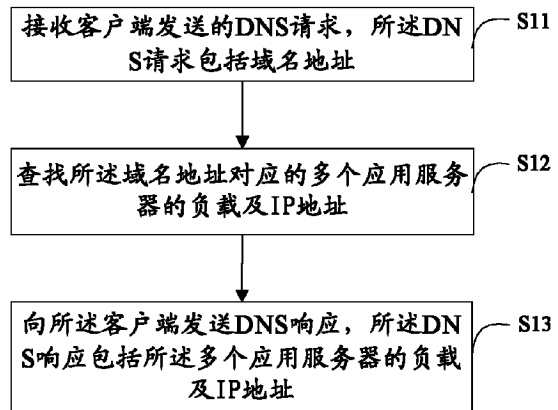
权利要求书 3 页 说明书 8 页 附图 3 页

(54) 发明名称

实现负载均衡的方法、DNS 服务器和客户端

(57) 摘要

本发明涉及通信技术领域,具体涉及实现负载均衡的方法、DNS 服务器和客户端,其中实现负载均衡的方法包括:接收客户端发送的 DNS 请求,该 DNS 请求包括域名地址;查找该域名地址对应的多个应用服务器的负载及 IP 地址;向该客户端发送 DNS 响应,该 DNS 响应包括上述多个应用服务器的负载及 IP 地址,以便于该客户端根据上述多个应用服务器的负载确定访问的 IP 地址。使用本发明,能够实现客户端根据多个应用服务器的负载情况访问负载较轻的应用服务器,以达到多个客户端均衡访问多个应用服务器的目的。



1. 一种实现负载均衡的方法,其特征在于,包括:  
接收客户端发送的 DNS 请求,所述 DNS 请求包括域名地址;  
查找所述域名地址对应的多个应用服务器的负载及 IP 地址;  
向所述客户端发送 DNS 响应,所述 DNS 响应包括所述多个应用服务器的负载及 IP 地址,以便于所述客户端根据所述多个应用服务器的负载确定访问的 IP 地址。
2. 如权利要求 1 所述的方法,其特征在于,所述接收客户端发送的 DNS 请求之前还包括:  
获取所述多个应用服务器的负载。
3. 如权利要求 2 所述的方法,其特征在于,所述获取所述多个应用服务器的负载包括:  
向所述多个应用服务器发送负载请求;  
接收所述多个应用服务器发送的负载响应,所述负载响应包括所述多个应用服务器的负载。
4. 如权利要求 3 所述的方法,其特征在于,所述负载响应还包括所述多个应用服务器的状态,所述多个应用服务器的状态包括可用或不可用;  
所述 DNS 响应还包括所述多个应用服务器的状态,以便于所述客户端根据所述多个应用服务器的负载及状态确定访问的 IP 地址。
5. 如权利要求 2 所述的方法,其特征在于,所述获取所述多个应用服务器的负载包括:  
接收所述多个应用服务器发送的负载消息,所述负载消息包括所述多个应用服务器的负载。
6. 如权利要求 5 所述的方法,其特征在于,所述接收所述多个应用服务器发送的负载消息包括:  
接收所述多个应用服务器周期性发送的负载消息。
7. 如权利要求 5-6 任一所述的方法,其特征在于,所述负载消息还包括所述多个应用服务器的状态,所述多个应用服务器的状态包括可用或不可用;  
所述 DNS 响应还包括所述多个应用服务器的状态,以便于所述客户端根据所述多个应用服务器的负载及状态确定访问的 IP 地址。
8. 一种实现负载均衡的方法,其特征在于,包括:  
向 DNS 服务器发送 DNS 请求,所述 DNS 请求包括域名地址;  
接收所述 DNS 服务器发送的 DNS 响应,所述 DNS 响应包括所述域名地址对应的多个应用服务器的负载及 IP 地址;  
根据所述多个应用服务器的负载确定访问的 IP 地址。
9. 如权利要求 8 所述的方法,其特征在于,所述根据所述多个应用服务器的负载确定访问的 IP 地址包括:  
将所述多个应用服务器中负载最轻的应用服务器对应的 IP 地址确定为访问的 IP 地址。
10. 如权利要求 8 所述的方法,其特征在于,所述 DNS 响应还包括所述多个应用服务器的状态,所述多个应用服务器的状态包括可用或不可用;  
所述根据所述多个应用服务器的负载确定访问的 IP 地址包括:  
根据所述多个应用服务器的负载及状态确定访问的 IP 地址。

11. 如权利要求 10 所述的方法,其特征在于,所述根据所述多个应用服务器的负载及状态确定访问的 IP 地址包括:

在状态为可用的应用服务器中,将负载最轻的应用服务器对应的 IP 地址确定为访问的 IP 地址。

12. 一种 DNS 服务器,其特征在于,包括:

接收单元,用于接收客户端发送的 DNS 请求,所述 DNS 请求包括域名地址;

查找单元,用于查找所述接收单元接收的 DNS 请求中域名地址对应的多个应用服务器的负载及 IP 地址;

发送单元,用于向所述客户端发送 DNS 响应,所述 DNS 响应包括所述查找单元查找出的多个应用服务器的负载及 IP 地址,以便于所述客户端根据所述多个应用服务器的负载确定访问的 IP 地址。

13. 如权利要求 12 所述的 DNS 服务器,其特征在于,还包括:

获取单元,用于在所述接收单元接收所述客户端发送的 DNS 请求之前,获取所述多个应用服务器的负载。

14. 如权利要求 13 所述的 DNS 服务器,所述获取单元获取所述多个应用服务器的负载包括:

向所述多个应用服务器发送负载请求;

接收所述多个应用服务器发送的负载响应,所述负载响应包括所述多个应用服务器的负载。

15. 如权利要求 13 所述的 DNS 服务器,所述获取单元获取所述多个应用服务器的负载包括:

接收所述多个应用服务器发送的负载消息,所述负载消息包括所述多个应用服务器的负载。

16. 如权利要求 15 所述的 DNS 服务器,所述接收所述多个应用服务器发送的负载消息包括:

周期性地接收所述多个应用服务器发送的负载消息。

17. 一种客户端,其特征在于,包括:

发送单元,用于向 DNS 服务器发送 DNS 请求,所述 DNS 请求包括域名地址;

接收单元,用于接收所述 DNS 服务器发送的 DNS 响应,所述 DNS 响应包括所述域名地址对应的多个应用服务器的负载及 IP 地址;

访问单元,用于根据所述多个应用服务器的负载确定访问的 IP 地址。

18. 如权利要求 17 所述的客户端,其特征在于,所述访问单元具体包括:

用于将所述多个应用服务器中负载最轻的应用服务器对应的 IP 地址确定为访问的 IP 地址。

19. 如权利要求 17 所述的客户端,其特征在于,所述 DNS 响应还包括所述多个应用服务器的状态,所述多个应用服务器的状态包括可用或不可用;

所述访问单元具体包括:

用于根据所述多个应用服务器的负载及状态确定访问的 IP 地址。

20. 如权利要求 19 所述的客户端,其特征在于,所述根据所述多个应用服务器的负载

及状态确定访问的 IP 地址包括：

在状态为可用的应用服务器中,将负载最轻的应用服务器对应的 IP 地址确定为访问的 IP 地址。

## 实现负载均衡的方法、DNS 服务器和客户端

### 技术领域

[0001] 本发明涉及通信技术领域,具体涉及实现负载均衡的方法、DNS 服务器和客户端。

### 背景技术

[0002] 域名解析系统 (DNS :Domain Name System) 负载均衡技术是最早的负载均衡解决方案, DNS 服务器为同一个域名地址配置多个不同的 IP 地址,因而当客户端访问域名地址时,能够获取 DNS 服务器根据该域名地址配置的 IP 地址列表,并且不同的客户端能够根据 IP 地址列表中不同的 IP 地址访问不同的应用服务器,达到访问应用服务器的负载均衡的目的。

[0003] DNS 负载均衡是一种简单而有效的实现负载均衡的方法,但是该种方案的缺点表现在不能区分不同的 IP 地址对应的应用服务器的差异,没有考虑这些应用服务器的负载能力,在没有考虑应用服务器的负载能力的情况下,简单的根据不同的 IP 地址访问不同的应用服务器,并没有实现客户端根据应用服务器的负载情况来均衡访问应用服务器,因此不是真正意义上的负载均衡。

### 发明内容

[0004] 本发明的实施例提供了实现负载均衡的方法、DNS 服务器和客户端,能够实现客户端根据应用服务器的负载情况来均衡访问应用服务器。

[0005] 本发明的实施例提供了一种实现负载均衡的方法,包括:接收客户端发送的 DNS 请求,所述 DNS 请求包括域名地址;查找所述域名地址对应的多个应用服务器的负载及 IP 地址;向所述客户端发送 DNS 响应,所述 DNS 响应包括所述多个应用服务器的负载及 IP 地址,以便于所述客户端根据所述多个应用服务器的负载确定访问的 IP 地址。

[0006] 本发明的另一个实施例提供了一种实现负载均衡的方法,包括:向 DNS 服务器发送 DNS 请求,所述 DNS 请求包括域名地址;接收所述 DNS 服务器发送的 DNS 响应,所述 DNS 响应包括所述域名地址对应的多个应用服务器的负载及 IP 地址;根据所述多个应用服务器的负载确定访问的 IP 地址。

[0007] 本发明的一个实施例提供了一种 DNS 服务器,包括:接收单元,用于接收客户端发送的 DNS 请求,所述 DNS 请求包括域名地址;查找单元,用于查找所述接收单元接收的 DNS 请求中域名地址对应的多个应用服务器的负载及 IP 地址;发送单元,用于向所述客户端发送 DNS 响应,所述 DNS 响应包括所述查找单元查找出的多个应用服务器的负载及 IP 地址,以便于所述客户端根据所述多个应用服务器的负载确定访问的 IP 地址。

[0008] 本发明一个实施例提供了一种客户端,包括:发送单元,用于向 DNS 服务器发送 DNS 请求,所述 DNS 请求包括域名地址;接收单元,用于接收所述 DNS 服务器发送的 DNS 响应,所述 DNS 响应包括所述域名地址对应的多个应用服务器的负载及 IP 地址;访问单元,用于根据所述多个应用服务器的负载确定访问的 IP 地址。

[0009] 从本发明实施例提供的以上技术方案可以看出,由于本发明实施例中客户端在向

DNS 服务器发送 DNS 请求后, DNS 服务器根据 DNS 请求中的域名地址查找对应的多个应用服务器的 IP 地址及负载, 以便于客户端能够根据多个应用服务器的负载情况确定合适的应用服务器作为访问对象, 实现多个客户端均衡访问多个应用服务器的目的。

### 附图说明

[0010] 为了更清楚地说明本发明实施例中的技术方案, 下面将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性的前提下, 还可以根据这些附图获得其他的附图。

[0011] 图 1 为本发明一个实施例提供的一种实现负载均衡的方法的流程图。

[0012] 图 2 为本发明另一个实施例提供的一种实现负载均衡的方法的流程图。

[0013] 图 3 为本发明一个实施例提供的一种实现负载均衡的方法的信令流程图。

[0014] 图 4 为本发明一个实施例提供的 DNS 查询报文格式示意图。

[0015] 图 5 为本发明一个实施例提供的扩展的 DNS 响应报文格式示意图。

[0016] 图 6 为本发明一个实施例提供的一种 DNS 服务器的结构图。

[0017] 图 7 为本发明一个实施例提供的一种客户端的结构图。

### 具体实施方式

[0018] 下面将结合本发明实施例中的附图, 对本发明实施例中的技术方案进行清楚、完整地描述。显然, 所描述的实施例仅仅是本发明一部分实施例, 而不是全部的实施例。基于本发明中的实施例, 本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例, 都属于本发明保护的范围。

[0019] 图 1 描述了本发明一个实施例提供的一种实现负载均衡的方法的流程, 该实施例包括:

[0020] S11、接收客户端发送的 DNS 请求, 所述 DNS 请求包括域名地址。

[0021] 具体的, DNS 服务器接收客户端发送的 DNS 请求, 该 DNS 请求中携带客户端访问的域名地址; 其中, DNS 服务器主要用于解析客户端访问的域名地址, 并将域名地址转换为对应的 IP 地址。

[0022] DNS 服务器在接收客户端发送的 DNS 请求之前, 会获取该域名地址对应的多个应用服务器的负载, 具体包括: DNS 服务器向该多个应用服务器发送负载请求, 并接收该多个应用服务器发送的负载响应, 该负载响应包括该多个应用服务器的负载, 进一步, 该负载响应还可以包括该多个应用服务器的状态, 包括可用或不可用; 可替代的, DNS 服务器接收该多个应用服务器主动上报的负载消息, 该负载消息包括该多个应用服务器的负载, 进一步, 还可以包括该多个应用服务器的状态, 包括可用或不可用, 进一步的, 该多个应用服务器可以周期性上报自身的负载消息。

[0023] S12、查找所述域名地址对应的多个应用服务器的负载及 IP 地址。

[0024] 具体的, DNS 服务器在接收客户端发送的 DNS 请求后, 获取其中携带的域名地址, 并根据该域名地址查找对应的多个应用服务器的 IP 地址和该多个应用服务器的负载。

[0025] 在本发明的实施例中, 应用服务器包括 Web 服务器、邮件服务器、FTP (File

Transfer Protocol) 服务器等,应用服务器的负载反映了应用服务器的承载能力,负载越大,表明应用服务器还能接受的访问越少,负载越小,表明应用服务器还能接受的访问越多。

[0026] S13、向所述客户端发送 DNS 响应,所述 DNS 响应包括所述多个应用服务器的负载及 IP 地址,以便于所述客户端根据所述多个应用服务器的负载确定访问的 IP 地址。

[0027] 具体的,DNS 服务器向客户端发送 DNS 响应消息,该 DNS 响应消息中携带上述查找到的多个应用服务器的负载及 IP 地址,以便于客户端根据该多个应用服务器的负载确定访问的 IP 地址。

[0028] 从上可知,本实施例中客户端在向 DNS 服务器发送 DNS 请求后,DNS 服务器根据 DNS 请求中的域名地址查找对应的多个服务器的负载及 IP 地址,以便于客户端能够根据多个应用服务器的负载情况确定合适的应用服务器作为访问对象,实现多个客户端均衡访问多个应用服务器的目的。

[0029] 图 2 描述了本发明另一个实施例提供的一种实现负载均衡的方法的流程,该实施例包括:

[0030] S21、向 DNS 服务器发送 DNS 请求,所述 DNS 请求包括域名地址。

[0031] 具体的,客户端向 DNS 服务器发送 DNS 请求,该 DNS 请求中携带客户端请求访问的域名地址。

[0032] S22、接收所述 DNS 服务器发送的 DNS 响应,所述 DNS 响应包括所述域名地址对应的多个应用服务器的负载及 IP 地址。

[0033] 具体的,客户端接收该 DNS 服务器发送的 DNS 响应,该 DNS 响应包括上述域名地址对应的多个应用服务器的负载及 IP 地址;其中,该多个应用服务器的负载及 IP 地址由该 DNS 服务器根据上述域名地址查找获取。

[0034] S23、根据所述多个应用服务器的负载确定访问的 IP 地址。

[0035] 具体的,客户端根据上述多个应用服务器的负载确定访问的 IP 地址,并根据该 IP 地址来访问获取相关资源。优选的,客户端将上述多个应用服务器中负载最轻的应用服务器对应的 IP 地址确定为访问的 IP 地址。可选的,上述 DNS 响应还可以包括多个应用服务器的状态,包括可用或不可用,此时客户端可以根据多个应用服务器的负载及状态确定访问的 IP 地址,比如,在状态为可用的应用服务器中,将负载最轻的应用服务器对应的 IP 地址确定为访问的 IP 地址。

[0036] 从上可知,本实施例中客户端在向 DNS 服务器发送 DNS 请求后,DNS 服务器根据 DNS 请求中的域名地址查找对应的多个服务器的负载及 IP 地址,以便于客户端能够根据多个应用服务器的负载情况确定合适的应用服务器作为访问对象,实现多个客户端均衡访问多个应用服务器的目的。

[0037] 图 3 描述了本发明一个实施例提供的一种实现负载均衡的方法的信令流程,需要说明的是,实际中 DNS 服务器会与多个应用服务器发生交互,图 3 中标示出的一个应用服务器不能理解成是对 DNS 服务器与多个应用服务器交互的限制,因此该实施例中应用服务器可以理解成是多个应用服务器。该实施例包括:

[0038] S31、发送负载请求。

[0039] DNS 服务器获取应用服务器的负载情况,具体的,DNS 服务器向应用服务器发送负

载请求,该负载请求可以是用户数据报协议 (UDP :User Datagram Protocol) 消息,并在该 UDP 消息中携带请求应用服务器负载的信息。比如,定义该 UDP 请求消息结构如下:

[0040]

UDP 首部	Data
--------	------

[0041] 其中,数据 Data 部分包括:

[0042]

字段名	类型	长度	描述
MessageType	Byte	2	UDP 请求消息

MessageType 表示该 UDP 消息用于请求应用服务器的负载;可选的,该 UDP 请求消息中还可以包括请求应用服务器状态的信息,此时 MessageType 表示该 UDP 消息用于请求应用服务器的负载和状态。

[0043] S32、返回负载响应。

[0044] 应用服务器在接收 DNS 服务器发送的负载请求后,向该 DNS 服务器返回负载响应,具体的,该负载响应可以是 UDP 响应消息,该 UDP 响应消息包含该应用服务器的负载;进一步,当 S31 中的 UDP 请求消息中还包含请求该应用服务器的状态信息时,则该 UDP 响应消息中还包含该应用服务器的状态,该状态包括可用或不可用。

[0045] DNS 服务器在获取应用服务器发送的负载响应后,根据负载响应中携带的应用服务器的负载,保存该应用服务器的 IP 地址及其负载的对应关系;进一步,如果该负载响应中还携带了该应用服务器的状态,还保存该应用服务器的 IP 地址及其状态的对应关系。

[0046] S311、发送负载消息。

[0047] 可替换的,DNS 服务器获取应用服务器的负载可以不采用 S31 和 S32 两步,而由应用服务器主动向 DNS 服务器发送自身的负载消息,从而使得 DNS 服务器获取应用服务器的负载。进一步,应用服务器可以周期性地向 DNS 服务器发送自身的负载消息;其中,该负载消息包括应用服务器的负载,进一步,该负载消息还可以包括应用服务器的状态,包括可用或不可用。

[0048] DNS 服务器在获取应用服务器发送的上述负载消息后,根据该负载消息中携带的应用服务器的负载,保存该应用服务器的 IP 地址及其负载的对应关系;进一步,如果该负载响应中还携带了该应用服务器的状态,还保存该应用服务器的 IP 地址及其状态的对应关系。

[0049] S33、发送 DNS 请求。

[0050] 客户端向 DNS 服务器发送 DNS 请求,该 DNS 请求中携带了客户端请求访问的域名地址。具体的,该 DNS 请求使用的是标准的 DNS 查询报文格式,如图 4 所示。其中,问题 401 表示 DNS 查询报文请求查询的内容,由域名 402、类型 403 和类 404 组成,域名 402 是客户端请求访问的域名地址,类型 403 表示该 DNS 报文请求查询的类型,比如由域名查询 IP 地址或由 IP 地址查询域名,类 404 通常为 1,表示因特网数据。

[0051] S34、查找域名地址对应的多个应用服务器的负载及 IP 地址。

[0052] DNS 服务器在接收了客户端发送的 DNS 请求后,根据该请求中携带的域名地址查



找对应的多个应用服务器的负载及 IP 地址。

[0053] 具体的,由于域名地址可以对应多个应用服务器,即多个 IP 地址,DNS 服务器根据该 DNS 请求中的域名地址查找出对应的多个应用服务器的 IP 地址;另外,由于 DNS 服务器中保存了 IP 地址及其对应的应用服务器的负载的对应关系,因此 DNS 服务器根据该域名地址可以查找出对应的多个应用服务器的负载及 IP 地址。可选的,DNS 服务器还可以保存 IP 地址及其对应的应用服务器的状态的对应关系,因此 DNS 服务器还可以查找出该域名地址对应的多个应用服务器的状态,此时要求 DNS 服务器在获取多个应用服务器的负载时,同时获取该多个应用服务器的状态。

[0054] S35、发送 DNS 响应。

[0055] DNS 服务器在查找出域名地址对应的多个应用服务器的负载及 IP 地址后,向客户端发送 DNS 响应,该 DNS 响应包括该多个应用服务器的负载及 IP 地址。

[0056] 具体的,该 DNS 响应使用扩展的 DNS 响应报文格式,如图 5 所示。其中,回答 501 表示该 DNS 响应报文响应的相关内容,在标准的 DNS 响应报文格式中,回答 501 由指针 502、类型 503、类 504、TTL505、长度 506 和 IP 地址 507 组成,其中,指针 502 表示客户端请求的域名地址;类型 503 和图 4DNS 查询报文中的类型 403 相同;类 504 通常是 1,表示因特网数据;TTL505 表示 DNS 响应报文的生存时间;长度 506 指资源数据(比如 IP 地址)的长度;IP 地址 507 表示查找出的客户端请求的域名地址对应的 IP 地址。

[0057] 本发明扩展的 DNS 响应报文中“回答”部分扩展的字段包括负载 508,进一步还可以包括状态 509,其中,负载 508 是指 IP 地址对应的应用服务器的负载情况,可以由具体的数值来反映负载的大小;状态 509 是指 IP 地址对应的应用服务器的状态,包括可用或不可用。

[0058] S36、发起访问。

[0059] 客户端在获取 DNS 服务器发送的 DNS 响应后,根据该 DNS 响应中携带的多个应用服务器的负载确定访问的 IP 地址,并向该确定访问的 IP 地址发起访问请求以获取相应的资源。优选的,客户端选择多个应用服务器中负载最轻的应用服务器对应的 IP 地址发起访问请求;进一步,如果 DNS 响应中还包含该多个应用服务器的状态,则客户端选择状态为可用的应用服务器中负载最轻的应用服务器发起访问,从而获取应用服务器的资源。

[0060] 因此,客户端根据应用服务器的负载情况来访问合适的应用服务器,可以有效的防止负载较重的应用服务器继续接受访问,而负载较轻的应用服务器却没有被访问,从而实现了合理的访问应用服务器,均衡了多个应用服务器的访问负载。

[0061] 从上可知,本实施例中客户端在向 DNS 服务器发送 DNS 请求后,DNS 服务器根据 DNS 请求中的域名地址查找对应的多个服务器的负载及 IP 地址,以便于客户端能够根据多个应用服务器的负载情况确定合适的应用服务器作为访问对象,实现多个客户端均衡访问多个应用服务器的目的。

[0062] 图 6 描述了本发明一个实施例提供的一种 DNS 服务器的结构,包括:

[0063] 接收单元 608,用于接收客户端发送的 DNS 请求,该 DNS 请求包括该客户端请求的域名地址。

[0064] 查找单元 609,用于根据接收单元 608 接收的 DNS 请求中的域名地址,查找出对应的多个应用服务器的负载及 IP 地址。

[0065] 发送单元 610,用于向该客户端发送 DNS 响应,该 DNS 响应包括查找单元 609 查找出的多个应用服务器的负载及 IP 地址,以便于该客户端根据该多个应用服务器的负载确定访问的 IP 地址。

[0066] 获取单元 611,用于在接收单元 608 接收客户端发送的 DNS 请求之前,获取上述多个应用服务器的负载,并保存该多个应用服务器的负载及 IP 地址的对应关系。因此,查找单元 609 根据接收单元 608 接收的域名地址和获取单元 611 获取的应用服务器的负载,查找出该域名地址对应的该多个应用服务器的负载。

[0067] 具体的,获取单元 611 可以通过向该多个应用服务器发送负载请求,再接收该多个应用服务器发送的负载响应来获取该多个应用服务器的负载情况,该过程与图 3 中 S31 和 S32 步骤相同,具体可以参见 S31 和 S32,此处不予赘述。

[0068] 可替换的,获取单元 611 可以通过直接接收该多个应用服务器上上报的负载消息来获取该多个应用服务器的负载,此时要求该多个应用服务器能够主动向 DNS 服务器发送自身的负载信息。进一步,获取单元 611 可以每隔一定的时间接收应用服务器发送的负载消息,即周期性的向 DNS 服务器发送负载消息,该过程与图 3 中的 S311 步骤相同,具体可以参见 S311,此处不予赘述。

[0069] 图 6 又描述了本发明一个实施例提供的一种 DNS 服务器的结构,该 DNS 服务器可以包括至少一个处理器 601,至少一个网络接口 604,存储器 605 和至少一个通信总线 602,该通信总线 602 用于实现这些组件之间的连接通信。该 DNS 服务器还可以包括用户接口 603,用于实现与用户交互。存储器 605 可以包括:

[0070] 操作系统 606,用于处理各种基础的系统服务和执行基于硬件的任务。

[0071] 应用单元 607,用于根据客户端发送的 DNS 请求,向该客户端返回携带多个应用服务器的负载及 IP 地址的 DNS 响应,具体的,应用单元 607 包括:

[0072] 接收单元 608,用于接收客户端发送的 DNS 请求,该 DNS 请求包括该客户端请求的域名地址。

[0073] 查找单元 609,用于根据接收单元 608 接收的 DNS 请求中的域名地址,查找出对应的多个应用服务器的负载及 IP 地址。

[0074] 发送单元 610,用于向该客户端发送 DNS 响应,该 DNS 响应包括查找单元 609 查找出的多个应用服务器的负载及 IP 地址,以便于该客户端根据该多个应用服务器的负载确定访问的 IP 地址。

[0075] 获取单元 611,用于在接收单元 608 接收客户端发送的 DNS 请求之前,获取上述多个应用服务器的负载,并保存该多个应用服务器的负载及 IP 地址的对应关系。因此,查找单元 609 根据接收单元 608 接收的域名地址和获取单元 611 获取的应用服务器的负载,查找出该域名地址对应的该多个应用服务器的负载。

[0076] 具体的,获取单元 611 可以通过向该多个应用服务器发送负载请求,再接收该多个应用服务器发送的负载响应来获取该多个应用服务器的负载情况,该过程与图 3 中 S31 和 S32 步骤相同,具体可以参见 S31 和 S32,此处不予赘述。

[0077] 可替换的,获取单元 611 可以通过直接接收该多个应用服务器上上报的负载消息来获取该多个应用服务器的负载,此时要求该多个应用服务器能够主动向 DNS 服务器发送自身的负载信息。进一步,获取单元 611 可以每隔一定的时间接收应用服务器发送的负载消

息,即周期性的向 DNS 服务器发送负载消息,该过程与图 3 中的 S311 步骤相同,具体可以参见 S311,此处不予赘述。

[0078] 上述 DNS 服务器执行了图 1 或图 3 所示的实施例的相关方法,具体的工作流程在此不再赘述。

[0079] 从上可知,本实施例中客户端在向 DNS 服务器发送 DNS 请求后,DNS 服务器根据 DNS 请求中的域名地址查找对应的多个服务器的负载及 IP 地址,以便于客户端能够根据多个应用服务器的负载情况确定合适的应用服务器作为访问对象,实现多个客户端均衡访问多个应用服务器的目的。

[0080] 图 7 描述了本发明一个实施例提供的一种客户端,该实施例包括:

[0081] 发送单元 708,用于向 DNS 服务器发送 DNS 请求,该 DNS 请求包括域名地址,即客户端请求访问的域名地址。

[0082] 接收单元 709,用于接收该 DNS 服务器返回的 DNS 响应,该 DNS 响应包括该域名地址对应的多个应用服务器的负载及 IP 地址。进一步,该 DNS 响应还可以包括多个应用服务器的状态,包括可用或不可用。具体的,接收单元 709 接收的 DNS 响应与图 3 中 S35 中的 DNS 响应相同,此处不予赘述。

[0083] 访问单元 710,用于根据接收单元 709 接收的多个应用服务器的负载确定访问的 IP 地址,从而访问相应的应用服务器。优选的,访问单元 710 可以向上述多个应用服务器中负载最轻的应用服务器的 IP 地址发起访问请求,进而获取应用服务器的资源。可选的,如果接收单元 709 接收的 DNS 响应中还包括多个应用服务器的状态,则访问单元 710 可以向状态为可用的应用服务器中负载最轻的应用服务器的 IP 地址发起访问请求。

[0084] 图 7 又描述了本发明一个实施例提供的一种装载了客户端的装置的结构,该装置可以包括至少一个处理器 701,至少一个网络接口 704,存储器 705 和至少一个通信总线 702,该通信总线 702 用于实现这些组件之间的连接通信。该装置还可以包括用户接口 703,用于实现与用户交互。存储器 705 可以包括:

[0085] 操作系统 706,用于处理各种基础的系统服务和执行基于硬件的任务。

[0086] 客户端 707,用于根据接收的 DNS 服务器发送的 DNS 响应中的多个应用服务器的负载确定访问的 IP 地址,具体的,客户端 707 包括:

[0087] 发送单元 708,用于向 DNS 服务器发送 DNS 请求,该 DNS 请求包括域名地址,即客户端请求访问的域名地址。

[0088] 接收单元 709,用于接收该 DNS 服务器返回的 DNS 响应,该 DNS 响应包括该域名地址对应的多个应用服务器的负载及 IP 地址。进一步,该 DNS 响应还可以包括多个应用服务器的状态,包括可用或不可用。具体的,接收单元 709 接收的 DNS 响应与图 3 中 S35 中的 DNS 响应相同,此处不予赘述。

[0089] 访问单元 710,用于根据接收单元 709 接收的多个应用服务器的负载确定访问的 IP 地址,从而访问相应的应用服务器。优选的,访问单元 710 可以向上述多个应用服务器中负载最轻的应用服务器的 IP 地址发起访问请求,进而获取应用服务器的资源。可选的,如果接收单元 709 接收的 DNS 响应中还包括多个应用服务器的状态,则访问单元 710 可以向状态为可用的应用服务器中负载最轻的应用服务器的 IP 地址发起访问请求。

[0090] 上述客户端执行了图 2 或图 3 所示的实施例的相关方法,具体的工作流程在此不

再赘述。

[0091] 从上可知,本实施例中客户端在向 DNS 服务器发送 DNS 请求后, DNS 服务器根据 DNS 请求中的域名地址查找对应的多个服务器的负载及 IP 地址,以便于客户端能够根据多个应用服务器的负载情况确定合适的应用服务器作为访问对象,实现多个客户端均衡访问多个应用服务器的目的。

[0092] 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流程,是可以通过计算机程序来指令相关的硬件来完成,上述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,上述的存储介质可为磁碟、光盘、只读存储记忆体 (ROM :Read-Only Memory) 或随机存储记忆体 (RAM :Random Access Memory) 等。

[0093] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的技术方案;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

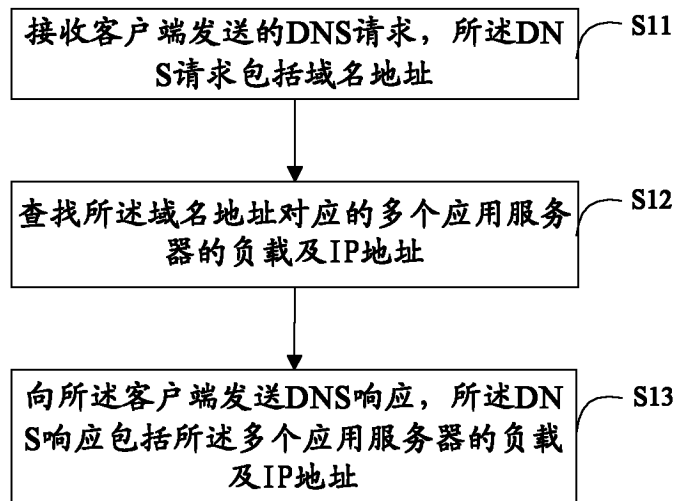


图 1

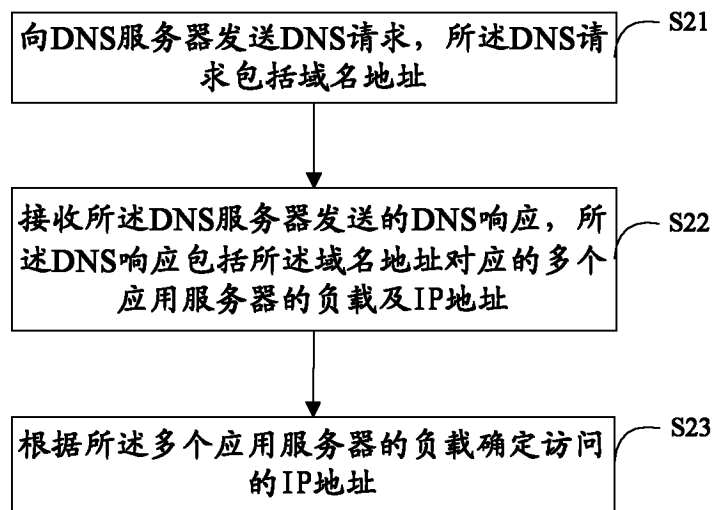


图 2

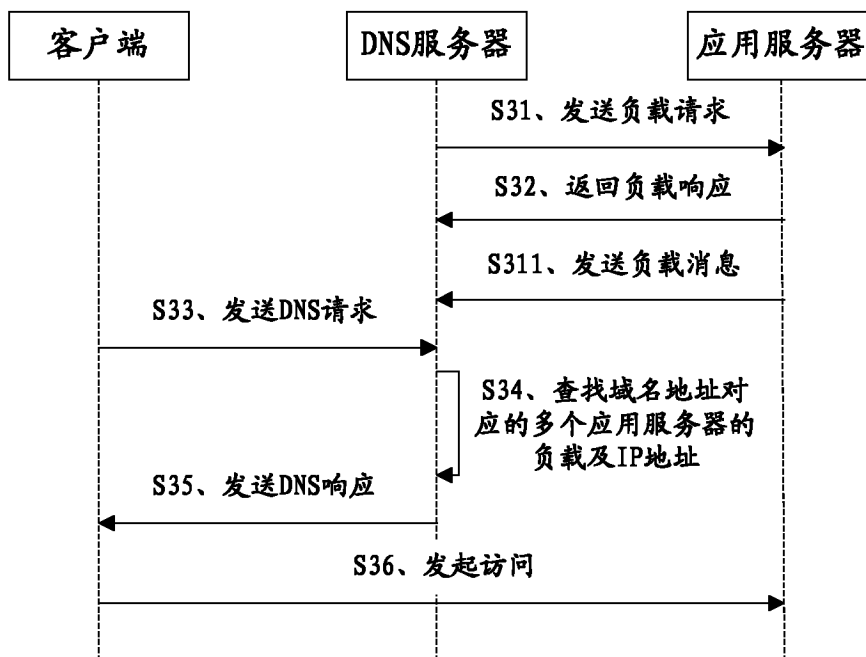


图 3

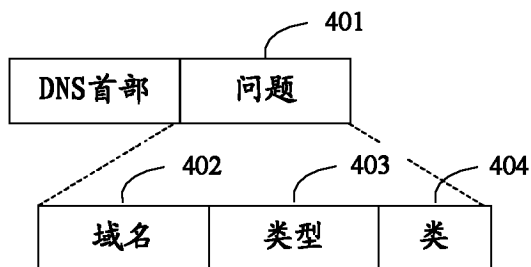


图 4

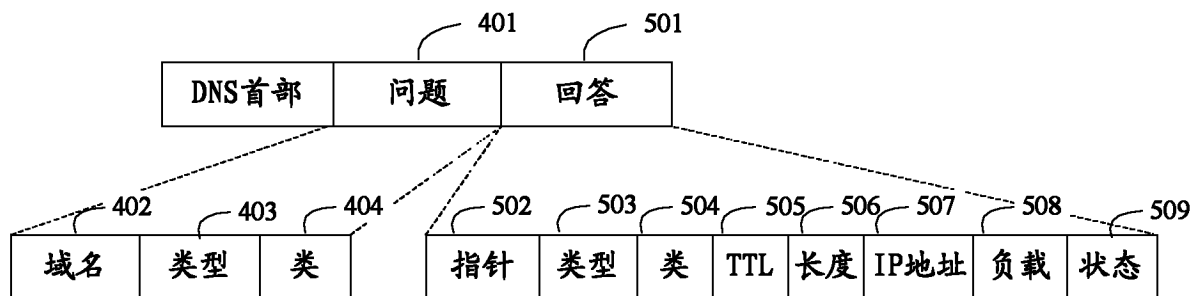


图 5

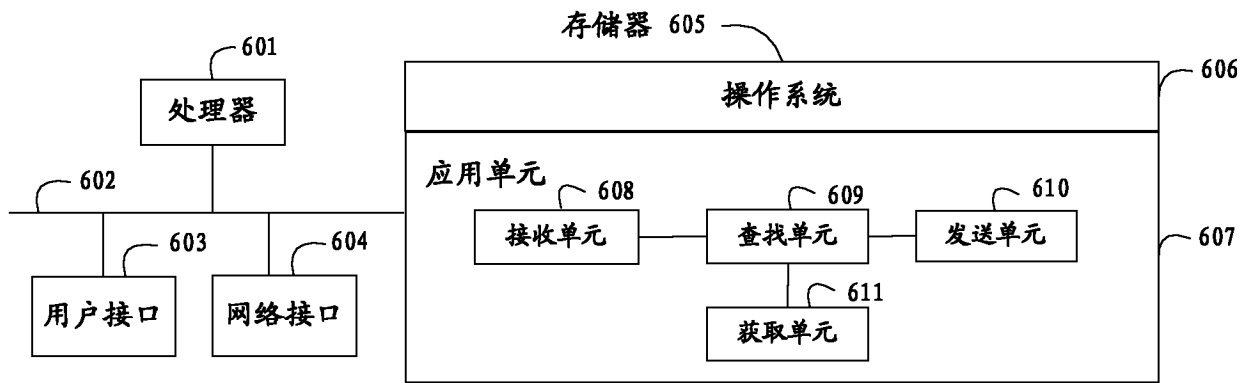


图 6

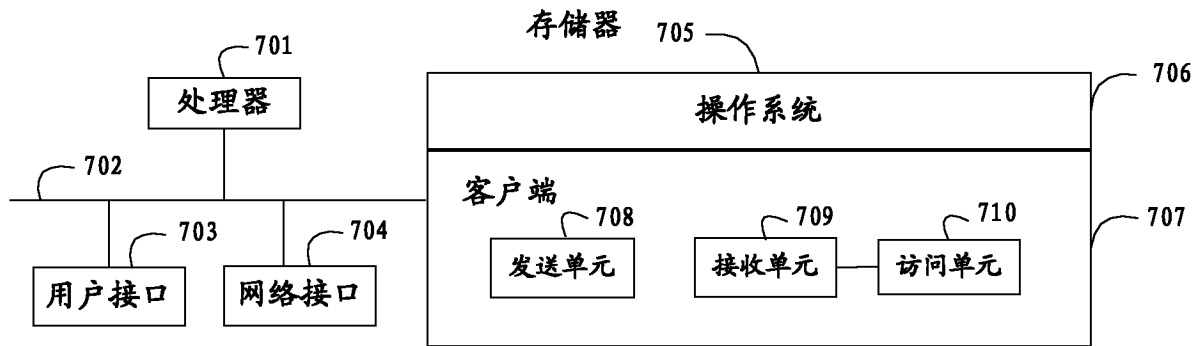


图 7