

一个模拟的负载均衡系统的实现

用户说明文档



V2.00

[pengzhengpu@foxmail.com](mailto:pengzhengpu@foxmail.com)

目录

[1. 系统概述 2](#_Toc363375989)

[2. 系统结构 2](#_Toc363375990)

[3. 分系统功能介绍 3](#_Toc363375991)

[3.1. 服务器(server.exe) 3](#_Toc363375992)

[3.2. 负载均衡器（LB.exe） 4](#_Toc363375993)

[3.2.1. 配置文件核对 4](#_Toc363375994)

[3.2.2. 启动LB.exe 6](#_Toc363375995)

[3.3. 客户端(client.exe) 7](#_Toc363375996)

[4. 系统使用流程 9](#_Toc363375997)

# 系统概述

负载均衡的常见实现方式大致如下：

将网络服务的地址（如公网IP地址、tcp套接字等）部署在负载均衡器上，而不是真实的服务器上；将负载均衡器作为网络服务的总入口，接收用户的所有访问请求；负载均衡器接收到用户的访问请求后，将访问请求按照一定的策略分发给某一台真实的服务器进行处理；真实服务器，对访问请求进行处理后，将处理结果发送给负载均衡器；负载均衡器接收到真实服务器的处理结果，将他发送给用户。

下图展示了一个负载均衡系统的组网结构，图中包含了1台负载均衡器，3台真实的服务器。公网IP配置在负载均衡器上，负载均衡器与真实服务器之间，则通过私网地址进行通讯。

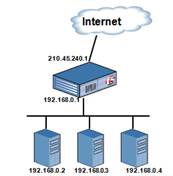


图 1 负载均衡组网图

# 系统结构

本系统包含如下3个可执行程序：

服务端(server.exe)――辅助程序，通过UDP端口，提供时间查询服务；

负载均衡器(LB.exe)――核心程序，用于实现负载均衡功能；

客户端(client.exe)――辅助程序，通过UDP端口，访问时间查询服务；

3个程序的协作关系如下图所示，其中客户端与服务端程序，需要起多个进程。

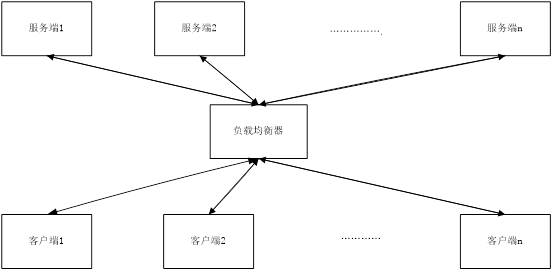


图 2系统协作

上图中，每一个方框表示一个进程。每一个进程拥有一个唯一的id（注意，这是由用户配置的id，并非操作系统为进程分配的pid），进程之间一律通过UDP协议进行通信。系统运行起来之后，客户端通过UDP协议向负载均衡器发送“时间请求”消息，负载均衡器通过UDP协议将消息分发给某个服务端进行处理。服务端返回“时间应答”消息给负载均衡器，负载均衡器将“时间应答”消息返回给客户端。

# 分系统功能介绍

## 服务器(server.exe)



图 3 server.exe软件界面—启动时

服务器软件（server.exe）启动界面如图 3所示：

1. 软件启动后，须先在左上角输入进程ID和UDP端口号，注意：发起多进程时进程ID和端口号不可重复，这里填写的信息须手动保存至负载均衡器的配置文件中；
2. 点击“开启服务”，即可接收负载均衡器分发的消息，并做出应答；

如图 4所示，为运行后的界面（调试-开）：

1. 服务器统计模块：可以显示收到消息条数（正确、错误），应答消息条数；
2. 调试开关：关闭时不显示消息详情，打开后可以显示；
3. 心跳报表开关：在调试开关打开的基础上，控制心跳信息是否需要在报表中显示；
4. 清除记录：在一次运行完毕后，可以选择“清除记录”，供下次测试使用；

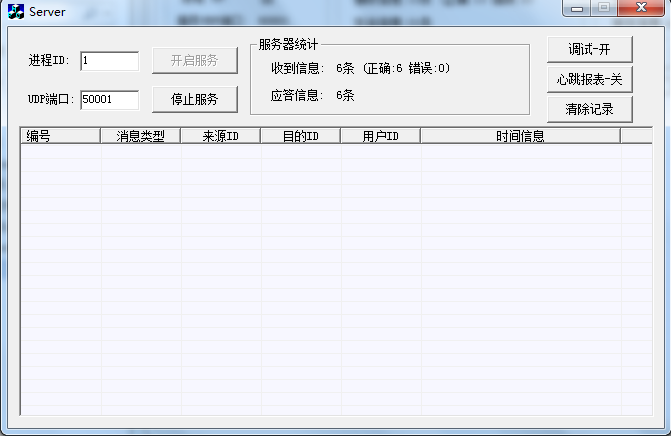


图 4 server.exe软件界面—运行后

## 负载均衡器（LB.exe）

### 配置文件核对

1. **LB配置信息：**

Process\_id：指定LB.exe的进程ID；

Client\_udp\_port：指定与客户端通信的UDP端口号；

Server\_udo\_port：指定与服务器通信的UDP端口号；

Server\_point\_number：共有服务器数量；

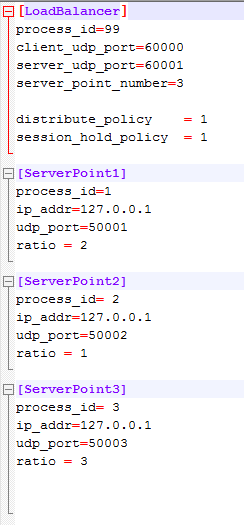
Distribute\_policy：负载均衡算法-请求信息分发策略,{0（轮转）,1（按比例权重）,2（按最快响应速度）}

Session\_hold\_policy：会话保持策略，{0（无会话保持功能），1（基于src\_id），2（基于usr\_id）}；

如图 5，选择的是按比例权重负载均衡算法和基于src\_id的会话保持。

1. **服务器配置信息：**

在启动LB.exe之前，请确认已启动服务器的进程ID、端口号、IP地址与配置文件中（文件名为”lbconfig.ini”）参数是否匹配。发起了多少个服务器进程，就在server\_point\_number处填相应的个数，每个服务器详细信息是否按照格式填入配置文件中。如**错误!未找到引用源。**所示，当前服务器个数为3，每个服务器参数如图填好。注意：各服务端信息在配置文件中的键名，须以[ServerPoint\*]的方式命令，且为顺序递增的，否则LB.exe无法识别该信息。



服务器配置信息

LB配置信息

图 5 配置文件介绍

**增加服务器方法:**

假设在3台服务器基础上增加一台，则将增加一台服务器的信息，更新到配置文件中：首先增加server\_point\_num的值，再新建一个键[ServerPoint4]，在其下增加process\_id=4; ip\_addr=127.0.0.1; udp\_port=5004;这些值可以视情况而取，但必须保证[ServerPoint\*]键名是按顺序递增的。

**减少服务器方法：**

假设减少2号服务器：首先减少server\_point\_num的值，再删掉二号服务器键名[ServerPoint2]和它所属的内容，最后将2号服务器后面服务器的键名都减少1。例如3号服务器键名更新为[ServerPoint2]，依次类推。

### 启动LB.exe

在添加好配置文件后，点击“LB.exe”即可启动程序，如图 6所示，软件界面上只能显示LB的配置信息，而不能更改这些信息。如要更改，须启动前在配置文件中更改。

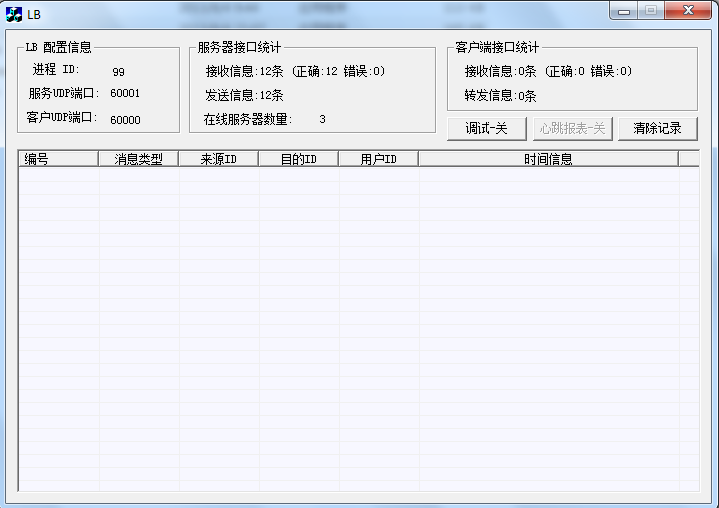


图 6 LB.exe软件界面—启动时

运行后的界面如图 7（调试-关）：服务器接口端的统计，可以统计与服务器交互过程中，发送的消息和接收的消息数量，其中包括接收正确数量和错误数量。还通过心跳健康检查显示当前在线可用服务器数量。客户端接口统计，可以统计与客户端程序交互过程，接收和转发的消息数，其中包括接收正确数量和错误数量。调试开关打开，可以在报表处显示消息详情。在测试完毕后，可以清除记录，供下次测试使用。

如图 7显示了，服务端接口发送和接受消息的条数，客户端接口发送和接受消息条数，“调试-开”在报表中显示心跳信息的详情，“心跳报表开”在报表中显示心跳信息。



图 7 LB.exe软件界面--运行后

在运行过程中，出现异常，会在写入到日志文件中（”LoadBalancer.log”）。日志详情按如下格式描述：[时间][事件描述][可能原因][错误码]。

## 客户端(client.exe)

双击client.exe，启动客户端程序如图 8所示，须填写服务器端的IP地址、端口号和进程ID，在负载均衡系统就是负载均衡器对应的配置参数。可以按需配置客户端的参数：进程ID，用户ID，以及一次性发送请求消息的条数。但请注意：发起多个客户端进程时，须保证用户ID不同。否则，LB.exe在转发服务器结果消息时，可能会出现错乱。

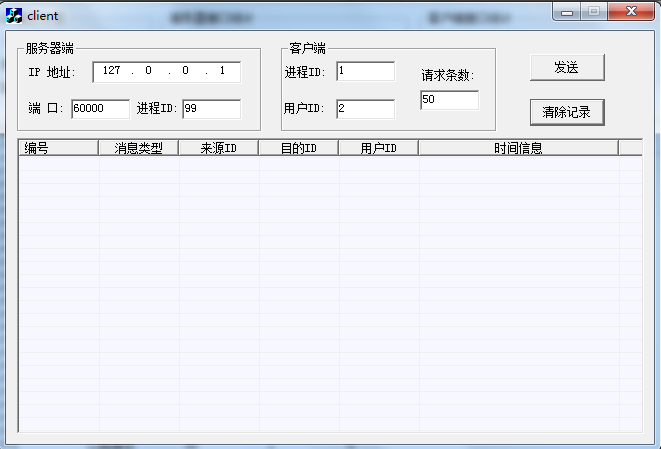


图 8 client.exe软件界面--启动时

点击“发送”后，软件界面如图 9所示，在运行过程，没有调试开关，直接显示发送和接收的消息详情。如果向一个不存在的服务器发送请求或者id号不匹配，则会引起短时间的停顿，然后发送下一个请求。在接收完毕后，会弹出一个对话款，显示发送和接收的消息条数，其中包括正确和错误的条数。当点击“确定”时，client.exe程序会退出。点击“取消”时，程序不退出，可以继续运行测试，不过建议在下次测试前，点击“清除记录”来清除之前运行的统计信息和消息报表显示的详情。

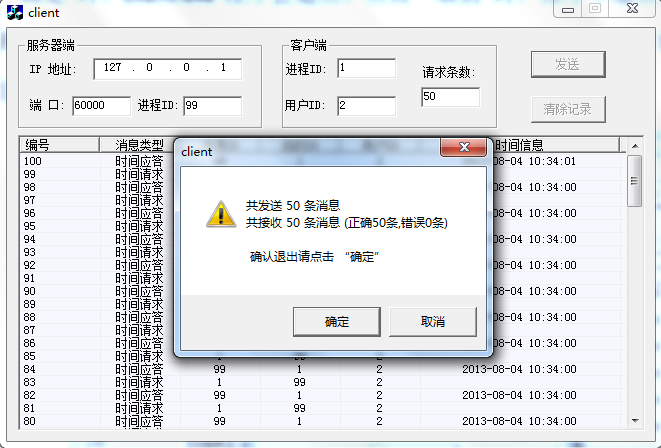


图 9 client.exe软件界面--运行后

# 系统使用流程

图 10 负载均衡系统使用流程图

可以按图 10流程执行测试，也可以先规划LB.exe配置文件服务器的参数，再依照此参数发起多个server.exe，但必须保证两者是统一的。否则，会导致系统异常。具体每一步骤可见“分系统功能介绍”一节。

在测试负载均衡算法和会话功能时，可在打开“调试-开”的基础上选择“心跳报表-关”方式，这样在报表中就不会显示心跳包的信息便于核对消息，但统计中仍会计算该信息。