# TCP长连接和短连接

### 什么是长连接、短连接

短连接：连接—>传输数据—>关闭；

长连接：连接—>传输数据—>传输数据🡪… 🡪关闭

长连接的连接数不能过多。

### 数据的发送接收方式

异步，对于每个连接，发送和接收数据分别有不同的线程负责

同步，对于每个连接，由同一个线程发送和接收。同步需要考虑超时问题。

单工，数据传输只支持数据在一个方向上传输

半双工，数据传输允许数据在两个方向上传输，但是，在某一时刻，只允许数据在一个方向上传输，它实际上是一种切换方向的单工通信；

全双工，数据通信允许数据同时在两个方向上传输，要求发送方和接收方都有独立的接收和发送能力。

单个Socket接收和发送可以同时进行，但接收不能同时，发送不能同时。

### 封包、粘包或半包、拆包

定义一个包的大小上限(8M)。（防止一个包过大，造成占用大量内存）。

注：UDP包可能丢失，乱序，但不存在粘包问题。TCP不存在丢包，乱序，但存在粘包问题。

粘包：一个包可能实际上分为了多块，或者多个包合成了一个包。

粘包的三种解决方案：

1. 固定长度包
2. 固定长度包头+可变长度包体（包体的长度在包头中定义）（如tlv报文格式 T和L字段长度固定， T字段表示报文类型（4byte）， L字段表示报文长度(16byte)、V字段表示报文的内容）注：T\V建议使用Encoding.UTF8编码，UTF8编码单元为字符，没有端的问题，L字段使用大端编码。 因为报文类型是一定的，如果是未识别的类型，则丢包。
3. 包头包尾标志（包头标志可以没有） （可尝试使用固定长度+包头包尾）

大包问题：传输的数据量较大，需要切割成多块发送。

对于UDP协议，单包最大长度为65535，其中包头长度20，所以单包最大长度最好不要超过60000.

网络是状态很好的局域网，那么UDP包分大点，提高系统的性能。不好，就分小于1464，这样可以减低丢包率。对于TCP来说，这个就要靠经验了，因为，TCP丢包可以自动重传，分大了，系统性能提高了，分包和错误重组可能会耗费时间，使传送时间延长，分小了，系统性能又降低了。大数据应该采用分包-封包处理，无论tcp还是udp建议一个包的长度控制在512字节左右。

1.**在链路层**，由以太网的物理特性决定了数据帧的长度为（46＋18）－（1500＋18），其中的18是数据帧的头和尾，也就是说**数据帧的内容最大为1500**（不包括帧头和帧尾），即MTU（Maximum Transmission Unit）为1500； 　  
2.**在网络层**，因为IP包的首部要占用20字节，所以这的MTU为1500－20＝1480；　  
3.**在传输层**，对于UDP包的首部要占用8字节，所以这的MTU为1480－8＝1472； 　　  
所以，在应用层，你的Data最大长度为1472。 （当我们的UDP包中的数据多于MTU(1472)时，发送方的IP层需要分片fragmentation进行传输，而在接收方IP层则需要进行数据报重组，由于UDP是不可靠的传输协议，如果分片丢失导致重组失败，将导致UDP数据包被丢弃）。 　　  
从上面的分析来看，在普通的局域网环境下，UDP的数据最大为1472字节最好（避免分片重组）。 　　  
但在网络编程中，Internet中的路由器可能有设置成不同的值（小于默认值），**Internet上的标准MTU值为576**，所以Internet的UDP/TCP编程时数据长度最好在576－20－8＝548字节以内。大数据应该采用分包-封包处理，无论tcp还是udp建议一个包的长度控制在512字节左右。

### 心跳包

心跳包一般是客户端定时通知服务器端我还在。服务器端一般也要定时轮询所有的连接，如果客户端在规定时间内没有响应，则认为服务器断线。

心跳包用于长连接保活和断线重连。心跳包周期一般可设置为16s。

### 连接的关闭

客户端发起关闭，服务器端receive返回0，然后服务器端就可以关闭连接了。

永远不要主动从服务器端关闭。服务器只有在认为客户端已经关闭或者连接中断时，才关闭连接，清理资源占用。

## Keepalive

Tcp中Keepalive是用来检测死链接的，TCP会在空闲了一段时间后发送数据给对方。KeepAliveTime 值控制 TCP/IP 尝试验证空闲连接是否完好的频率。如果这段时间内没有活动，则会发送保持活动信号。如果网络工作正常，而且接收方是活动的，它就会响应。如果没有收到 keep-alive 应答，keep-alive请求 将在每 KeepAliveInterval 秒重发一次。

### 重连

客户端如果连接或者发送到服务器数据出错或者其它认为客户端已经断开的情况，客户端释放资源并重新创建链接。

### 服务器设计

服务器在不能为每个连接分配一个线程。而是构建多个自定义连接池的方式。每个连接池包含一个接收数据线程，一个数据处理线程和一个数据发送线程。每个连接池同时包含两个队列：一个发送队列和一个接受队列。线程池的数目通常设定为factor\*cpu核数+2。Factor是一个>=1.0的常数因子，通常factor=2.0。接收数据线程扫描连接池中所有连接，如果某个连接有数据，则从该链接中获取数据并解析数据，并放到接受队列中。数据处理线程扫描接收队列，并处理接收到的数据，如果要向客户端发送数据，则把要发送的数据放到发送队列。发送线程扫描发送队列将数据发送到客户端。目前，三个线程处理任务的cpu比重是不一样的。比如数据发送线程，仅仅负责发送数据。**在未来，可以使多个连接池共用数据发送线程。**

如果是短连接，可以考虑使用线程池。

（多路IO方案，如果是长连接，并且可以接受单工通信，可以通过BeginReceive实现通信，一个线程接受连接，放到队列中，另一个线程从队列中获取连接，并BeginReceive调用，）。

private void AcceptCallback(IAsyncResult result)

{

ConnectionInfo connection = new ConnectionInfo();

try

{

// Finish Accept

Socket s = (Socket)result.AsyncState;

connection.Socket = s.EndAccept(result);

connection.Buffer = new byte[255];

lock (\_connections) \_connections.Add(connection);

// Start Receive and a new Accept

connection.Socket.BeginReceive(connection.Buffer, 0,

connection.Buffer.Length, SocketFlags.None,

new AsyncCallback(ReceiveCallback), connection);

\_serverSocket.BeginAccept(new AsyncCallback(AcceptCallback),

result.AsyncState);

}

catch (SocketException exc)

{

CloseConnection(connection);

Console.WriteLine("Socket exception: " + exc.SocketErrorCode);

}

catch (Exception exc)

{

CloseConnection(connection);

Console.WriteLine("Exception: " + exc);

}

}

private void ReceiveCallback(IAsyncResult result)

{

ConnectionInfo connection = (ConnectionInfo)result.AsyncState;

try

{

int bytesRead = connection.Socket.EndReceive(result);

if (0 != bytesRead)

{

lock (\_connections)

{

foreach (ConnectionInfo conn in \_connections)

{

if (connection != conn)

{

conn.Socket.Send(connection.Buffer, bytesRead,

SocketFlags.None);

}

}

}

connection.Socket.BeginReceive(connection.Buffer, 0,

connection.Buffer.Length, SocketFlags.None,

new AsyncCallback(ReceiveCallback), connection);

}

else CloseConnection(connection);

}

catch (SocketException exc)

{

CloseConnection(connection);

Console.WriteLine("Socket exception: " + exc.SocketErrorCode);

}

catch (Exception exc)

{

CloseConnection(connection);

Console.WriteLine("Exception: " + exc);

}

}

新连接到来时，新链接选择连接数最小的连接池。从而平衡各个线程的处理负载。这是基于每个连接数据交换的量是接近的合理假设下的。

通过以上架构，可以看出服务器采用tcp双工通信。

每个请求应该明确自己想要的东西，每个响应应该明确自己的回答。请求之间不应该有上下文关联。请求和响应之间也不应该有上下文关联。

在性能要求高的情况下，可采用protobuf或者json进行序列化和反序列化。一般情况下，采用XmlSerializer或者BinaryFormatter来进行序列化或者反序列化。

负载均衡可以采用主从服务器实现。客户端先向服务器请求可用的从服务器。从服务器的选择按照服务器当前连接最少。主服务器不应该或者应该较少的处理业务请求，主服务器的主要任务是协调各个从服务器和将连接分配到从服务器。

对于快速创建、快速销毁的内存尽量使用结构，而不是类。

### 三次握手、四次握手

### 大端、小端、网络字节序、主机字节序

网络字节序是TCP/IP协议规定好的数据表示格式，与具体的硬件、操作系统无关。网络字节序使用Big endian模式。

主机字节序有具体的cpu决定，可以是大端，也可以是小端。

网络通讯中，发送方一般将数据转换为网络字节序发送，接收方在转换为本地字节序解析。

UTF8使用1-4个字节来表示，单字节编码的第一字节为[00-7F]，双字节编码的第一字节为[C2-DF]，三字节编码的第一字节为[E0-EF]。这样只要看到第一个字节的范围就可以知道编码的字节数。 建议使用Encoding.UTF8编码，UTF8编码单元为字符，没有端的问题。UTF8看第一个字节就知道如何解释了。所以UTF-8不需要Byte Order Mark. UTF-8易于传输，一方面是因为它省空间。另一方面是它容易识别和生成，没有端的问题。

### 负载均衡

Socket负载均衡和Http负载均衡

LVS\ngnix\apache\客户端自己实现选择

如果事先知道会有多少个连接，在服务器进行选择，做主从服务器。主服务器只有一个，分服务器有多个，分服务器向主服务器进行注册，客户端向主服务器请求可用的服务器，然后断开和主服务器的连接，再和所请求到的服务器进行通信。

长连接当连接数较多时，不能一个连接一个线程。

### 分布式缓存

Memched\redis

事务

保存数据时，先清缓存，再保存，保存后，再清缓存。查询数据时，先查缓存，再查数据库。

### Memcached

Memcached是免费、开源、高性能、分布式内存对象缓存系统，它通过缓存对象来减少对数据库的访问。

### Broadcast MultiCast 广播和多播

只有UDP支持广播和多播。广播和多播只占用一份网络。广播使用地址255.255.255.255，广播将消息发送到同一广播网络中，所有的主机都能接收广播消息，广播消息不会被路由器转发。广播一般用于局域网游戏中交换消息。

地址(224.0.0.0 to 239.255.255.255)是用于多播的。在多播中需要设置TTl值（Time to live），每一个ip数据报文中都包含一个TTL，每当有路由器转发该报文时，TTL减1，知道减为0时，生命周期结束，报文即时没有到达目的地，也立即宣布死亡。多播将消息发送到特定的地址，对该地址注册为有兴趣的host将会接收到该消息。客户端需要接收消息，将注册到该地址，当客户端Close的时候自动注销关注。

### 杂项

无状态设计，即每个响应不必预先知道请求的信息。可以独立处理响应。

new IPEndPoint(IPAddress.Any, 0)// creates an IPEndpoint using any available IP address on the local computer, and any available port number.

IPAddress.Any表示监听所有活动的网络接口

端口0表示所有可用的端口

0FEH：如果一个字节以A-F开始，前面加0，H表示16进制数

1. 不定长消息标志消息和Field边界的方法有：
   * 1. 分界符  每个消息字段和消息结尾加分界符，分界符不能出现在消息本身中
     2. 显式指定消息字段和消息本身的长度。
2. 字符串分次编码还是一次编码，对最终结果没有影响。 假设E()表示编码函数—将字符串映射为byte数组。 假设a,b为任意字符串。E(a)表示a的字节编码序列。应该选用具有E(a+b) = E(a)+E(b)特性的编码方式，如：[UTF8Encoding.UTF8Encoding()](https://msdn.microsoft.com/zh-cn/library/s756abs9(v=vs.110).aspx)。Encoding.UTF8使用了BOM,因此不满足特性。经过测试Encoding.UTF8在getBytes(s)时没有加BOM.

byte-order indicator (the 2-byte sequence 254–255 for big-endian, and 255–254 for little-endian)

1. Utf-8 是 Unicode 编码，表示为一至四个字节序列的每个码位。与 utf-16 和 utf-32 编码的 utf-8 编码不需要"字节排序方式"；编码方案不需要考虑处理器是否-big-endian 或 little-endian 格式相同。
   1. 您可以实例化 UTF8Encoding 中通过多种方式，具体取决于是否要向其提供字节顺序标记 (BOM) 以及您是否想要启用错误检测的对象。下表列出了构造函数和 [Encoding](https://msdn.microsoft.com/zh-cn/library/system.text.encoding(v=vs.110).aspx) 返回的属性 UTF8Encoding 对象。如果不使用错误检测不会引发异常，并通常忽略无效的序列。经过测试Encoding.UTF8在getBytes(s)时没有加BOM.

|  |  |  |
| --- | --- | --- |
| **编号** | **BOM** | **错误检测** |
| [Encoding.UTF8](https://msdn.microsoft.com/zh-cn/library/system.text.encoding.utf8(v=vs.110).aspx) | 是 | 无 （替换回退） |
| [UTF8Encoding.UTF8Encoding()](https://msdn.microsoft.com/zh-cn/library/s756abs9(v=vs.110).aspx) | 否 | 无 （替换回退） |
| [UTF8Encoding.UTF8Encoding(Boolean)](https://msdn.microsoft.com/zh-cn/library/s064f8w2(v=vs.110).aspx) | 可配置 | 无 （替换回退） |
| [UTF8Encoding.UTF8Encoding(Boolean, Boolean)](https://msdn.microsoft.com/zh-cn/library/302sbf78(v=vs.110).aspx) | 可配置 | 可配置 |

|  |  |  |
| --- | --- | --- |
| I/O Operation | Socket Type | Blocking Avoidance Options |
| Accepting a  new connection | Socket | 1. Set the socket to nonblocking before calling  Accept(). |

2. Call Poll() or Select() on the socket before calling

Accept().

TcpListener 1. Only call AcceptSocket() or AcceptTcpClient() if

Pending() returns true.

|  |  |  |
| --- | --- | --- |
| Making a | Socket | 1. Set the socket to nonblocking before calling |
| new connection |  | Connect(). |

2. Call Poll() or Select() on the socket before calling

Connect().

Send Socket 1. Set the socket to nonblocking before calling Send() or

SendTo().

2. Call Poll() or Select() on the socket before calling

Send() or SendTo().

3. Set the SendTimeout socket option before calling Send()

or SendTo().

TcpClient 1. Set the SendTimeout property before calling Write() on the network stream.

Receive Socket 1. Set the socket to nonblocking before calling Receive()

or ReceiveFrom().

2. Call Poll() or Select() on the socket before calling

Receive() or ReceiveFrom().

3. Set the ReceiveTimeout socket option before calling

Receive() or ReceiveFrom().

4. Only call Receive() or ReceiveFrom() if property

Available > 0.

TcpClient 1. Set the ReceiveTimeout property before calling Read()

on the network stream.

2. Only call Read() on the TcpClient’s network stream if the DataAvailable property is true. (The Length property is not supported for NetworkStream.)

Shutdown(SocketShutdown.Send)后，客户端receive将会返回0.关闭发送后，还可以接收数据。

因为TCP是面向连接的，保证了不丢消息，所以如果接受者的recvQ不及时被读取，发送者的sendQ将被堵塞，send将发送不出消息，被阻塞掉。因此不要两边同时写，除非写和读放到不同的线程中。

   
Socket的关闭，当Socket调用Close()或者Shutdown(SocketShutdown.Send)时，底层tcp首先将在SendQ里剩余的消息发送出去，然后发送一个closing TCP握手消息，告诉接收TCP RecvQ中没有更多消息。（握手消息到达tcp协议层，不会到达应用程序层，但会在Read()中返回0.）然后TCP等待暗示数据都发送到RecvQ的握手消息，一旦接收到该消息，Socket被认为是Half Closed,然后对方发起一个对称的关闭请求。  Close()在Closing握手消息发送后，返回。因为Close是在发送握手消息之后返回的，实际上有可能SendQ的消息并没有完全发送到接收端，一个更合理的方式是实现应用程序级别的Close回话。

 Socket提供了Linger来控制Close()等待握手完成的时间，Close将会Block，直到Closing握手完成或者超时。socket.SetSocketOption(SocketOptionLevel.Socket,SocketOptionName.Linger,(object) new LingerOption(true,10));

1. 多路复用：同一台机器上的不同Socket可以有相同的ip和端口。数据包和socket的对应关系，采用如下原则：
   1. socket的接口必须与packet的目标接口一致
   2. 如果socket的地址为IPAddess.Any  即（\*）被认为是匹配的
   3. 如果多于一个Socket的地址与packet的目标地址匹配，the one that matches using the fewest \* wildcards gets the packet.
2. socket.SetSocketOption(SocketOptionLevel.Socket, SocketOptionName.ReuseAddress,1)