

为什么说基于TCP的移动端IM仍然需要心跳保活？



来自: Loly 2016-09-06 17:24:07

1、前言

很多人认为，TCP协议自身先天就有KeepAlive机制，为何基于它的通讯链接，仍然需要在应用层实现额外的心跳保活？本文将从移动端IM实践的角度告诉你，即使使用的是TCP协议，应用层的心跳保活仍旧必不可少。

有关TCP协议的权威理论介绍，请参见《TCP/IP详解》这本书。

说明：本文引用了网易云信项望峰的技术文章，感谢分享。（本文同步发布于：<http://www.52im.net/thread-33-1-1.html>）

2、学习交流

- 即时通讯开发交流群：215891622 [推荐]

- 移动端IM开发推荐文章：《新手入门一篇就够：从零开发移动端IM》

3、参考资料

- 《TCP/IP详解 - 第11章·UDP：用户数据报协议》
- 《TCP/IP详解 - 第17章·TCP：传输控制协议》
- 《TCP/IP详解 - 第18章·TCP连接的建立与终止》
- 《TCP/IP详解 - 第21章·TCP的超时与重传》
- 《通俗易懂·深入理解TCP协议（上）：理论基础》
- 《通俗易懂·深入理解TCP协议（下）：RTT、滑动窗口、拥塞处理》
- 《理论经典：TCP协议的3次握手与4次挥手过程详解》
- 《计算机网络通讯协议关系图（中文珍藏版）》
- 《NAT详解：基本原理、穿越技术(P2P打洞)、端口老化等》

4、本文源起

做移动端IM多年以来，经常会与相关人员进行讨论和交流。也经常会碰到些较真的技术人员询问技术细节，如主流的移动端IM如何做心跳、如何保证消息必达、如何加快文件上传等。因为平时工作太忙，没有时间深入整理和总结，往往只能简略介绍，并不能具体展开，于是决定写成文字，也有了有关移动IM问题处理的系列文章。

5、什么是心跳保活？

为什么说基于TCP的移动端IM仍然需要心跳保活？_a.jpg

在使用 TCP 长连接的 IM 服务设计中，往往都会涉及到心跳。心跳一般是指某端(绝大多数情况下是客户端)每隔一定时间向对端发送自定义指令，以判断双方是否存活，因其按照一定间隔发送，类似于心跳，故被称为心跳指令。

有兴趣了解IM/推送的心跳保活技术的文章，请参见：

- 《Android进程保活详解：一篇文章解决你的所有疑问》
- 《Android端消息推送总结：实现原理、心跳保活、遇到的问题等》
- 《微信团队原创分享：Android版微信后台保活实战分享(进程保活篇)》
- 《微信团队原创分享：Android版微信后台保活实战分享(网络保活篇)》
- 《移动端IM实践：实现Android版微信的智能心跳机制》
- 《移动端IM实践：WhatsApp、Line、微信的心跳策略分析》

>> 更多同类文章

6、TCP协议不是自带KeepAlive的吗？

那么问题就随之而来了：为什么需要在应用层做心跳，难道 TCP 不是个可靠连接吗？我们不能够依赖 TCP 做断线检测吗？比如使用 TCP 的 KeepAlive 机制来实现。应用层心跳是目前的最佳实践吗？怎样的心跳才是最佳实践。



即时通讯技术开发

2 人聚集在这个小组

加入小组



最新话题 (更多)

- 理论经典：TCP协议的3次握手与4次挥手过程详解 (Loly)
- 技术往事：改变世界的TCP/IP协议（珍贵多图、手机... (Loly)
- 为什么说基于TCP的移动端IM仍然需要心跳保活？ (Loly)
- 简述移动端IM开发的那些坑：架构设计、通信协议和... (Loly)
- 移动端IM系统的协议选型：UDP还是TCP？ (Loly)





很多做移动端IM的同行，以前确实没有仔细考虑过这些问题，潜意识里想当然的认为这仅仅只是个简单的心跳而已啊。好吧，事实并非这么简单，请继续往下看。

在这里发现跟你一样特别的人，并与之交流...

注册

登录

第三方登录:  



下，每一次请求都只是简单的数据发送和接受，免去了 DNS 解析，连接建立等时间，大大加快了请求的速度，同时也有利于接受服务器的实时消息。但前提是连接可用。

如果连接无法很好地保持，每次请求就会变成撞大运：运气好，通过长连接发送请求并收到反馈。运气差，当前连接已失效，请求迟迟没有收到反馈直到超时，又需要一次连接建立的过程，其效率甚至还不如 HTTP。而连接保持的前提必然是检测连接的可用性，并在连接不可用时主动放弃当前连接并建立新的连接。

基于这个前提，必须要有一种机制用于检测连接可用性。同时移动网络的特殊性也要求客户端需要在空余时间发送一定的信令，避免连接被回收。详见微信和运营商的撕B（另一篇针对微信的信令风暴技术研究文章请见：《微信对网络影响的技术试验及分析》）。

而对于服务器而言，能够及时获悉连接可用性也非常重要：一方面服务器需要及时清理无效连接以减轻负载，另一方面也是业务的需求，如游戏副本中服务器需要及时处理玩家掉线带来的问题。

8、TCP的KeepAlive无法替代应用层心跳保活机制的原因

上面说了保持连接的重要性，那么现在回到具体实现上。为什么我们需要使用应用层心跳来做检测，而不是直接使用 TCP 的特性呢？

我们知道 TCP 是一个基于连接的协议，其连接状态是由一个状态机进行维护，连接完毕后，双方都会处于 **established** 状态，这之后的状态并不会主动进行变化。这意味着如果上层不进行任何调用，一直使 TCP 连接空闲，那么这个连接虽然没有任何数据，但仍是保持连接状态，一天、一星期、甚至一个月，即使在这期间中间路由崩溃重启无数次。举个现实中经常遇到的栗子：当我们 ssh 到自己的 VPS 上，然后不小心踢掉网线，此时的网络变化并不会被 TCP 检测出，当我们重新插回网线，仍旧可以正常使用 ssh，同时此时并没有发生任何 TCP 的重连。

有人会说 TCP 不是有 KeepAlive 机制么，通过这个机制来实现不就可以了吗？但是事实上，TCP KeepAlive 的机制其实并不适用于此。Keep Alive 机制开启后，TCP 层将在定时时间到后发送相应的 KeepAlive 探针以确定连接可用性。一般时间为 7200 s（详情请参见《TCP/IP详解》中第23章），失败后重试 10 次，每次超时时间 75 s。显然默认值无法满足我们的需求，而修改过设置后就可以满足了吗？答案仍旧是否定的。

因为 TCP KeepAlive 是用于检测连接的死活，而心跳机制则附带一个额外的功能：检测通讯双方的存活状态。两者听起来似乎是一个意思，但实际上却大相径庭。

考虑一种情况，某台服务器因为某些原因导致负载超高，CPU 100%，无法响应任何业务请求，但是使用 TCP 探针则仍旧能够确定连接状态，这就是典型的连接活着但业务提供方已死的状态，对客户端而言，这时的最好选择就是断线后重新连接其他服务器，而不是一直认为当前服务器是可用状态，一直向当前服务器发送些必然会失败的请求。

从上面我们可以知道，KeepAlive 并不适用于检测双方存活的场景，这种场景还得依赖于应用层的心跳。应用层心跳有着更大的灵活性，可以控制检测时机，间隔和处理流程，甚至可以在心跳包上附带额外信息。从这个角度而言，应用层的心跳的确是最佳实践。

9、心跳保活机制的实现方案参考

从上面我们可以得出结论，目前而言，应用层心跳的确是检测连接有效性，双方是否存活的最佳实践，那么剩下的问题就是如何实现。

最简单粗暴做法当然是定时心跳，如每隔 30 秒心跳一次，15 秒内没有收到心跳回包则认为当前连接已失效，断开连接并进行重连。这种做法最直接，实现也简单。唯一的问题是耗电和耗流量。以一个协议包 5 个字节计算，一天收发 2880 个心跳包，一个月就是 $5 * 2 * 2880 * 30 = 0.8 \text{ M}$ 的流量，如果手机上多装几个 IM 软件，每个月光心跳就好几兆流量没了，更不用说频繁的心跳带来的电量损耗。

既然频繁心跳会带来耗电和耗流量的弊端，改进的方向自然是减少心跳频率，但也不能过于影响连接检测的实时性。基于这个需求，一般可以将心跳间隔根据程序状态进行调整，当程序在后台时(这里主要考虑安卓)，尽量拉长心跳间隔，5 分钟、甚至 10 分钟都可以。

而当 App 在前台时则按照原来规则操作。连接可靠性的判断也可以放宽，避免一次心跳超时就认为连接无效的情况，使用错误积累，只在心跳超时 n 次后才判定当前连接不可用。当然还有一些小 **trick** 比如从收到的最后一个指令包进行心跳包周期计时而不是固定时间，这样也能够一定程度减少心跳次数。

（本文同步发布于：<http://www.52im.net/thread-33-1-1.html>）

附录：更多IM技术文章

[1] 网络编程基础资料：

《TCP/IP详解 卷1：协议（<http://www.52im.net/topic-tcpipvol1.html>）》：第11章·UDP：用户数据报协议（<http://docs.52im.net/extend/docs/book/tcpip/vol1/11/>）、第17章·TCP：传输控制协议（<http://docs.52im.net/extend/docs/book/tcpip/vol1/17/>）、第18章·TCP连接的建立与终止（<http://docs.52im.net/extend/docs/book/tcpip/vol1/18/>）、第21章·TCP的超时与重传（<http://docs.52im.net/extend/docs/book/tcpip/vol1/21/>）

《[通俗易懂]深入理解TCP协议（上）：理论基础》<http://www.52im.net/thread-513-1-1.html>

《[通俗易懂]深入理解TCP协议（下）：RTT、滑动窗口、拥塞处理》<http://www.52im.net/thread-515-1-1.html>

《理论经典：TCP协议的3次握手与4次挥手过程详解》<http://www.52im.net/thread-258-1-1.html>

《理论联系实际：Wireshark抓包分析TCP 3次握手、4次挥手过程》<http://www.52im.net/thread-275-1-1.html>

《计算机网络通讯协议关系图（中文珍藏版）》<http://www.52im.net/thread-180-1-1.html>

《NAT详解：基本原理、穿越技术(P2P打洞)、端口老化等》<http://www.52im.net/thread-50-1-1.html>

《UDP中一个包的大小最大能多大？》<http://www.52im.net/thread-29-1-1.html>

《Java新一代网络编程模型AIO原理及Linux系统AIO介绍》<http://www.52im.net/thread-306-1-1.html>

《NIO框架入门(三)：iOS与MINA2、Netty4的跨平台UDP双向通信实战》<http://www.52im.net/thread-378-1-1.html>

《NIO框架入门(四)：Android与MINA2、Netty4的跨平台UDP双向通信实战》<http://www.52im.net/thread-388-1-1.html>

>> 更多同类文章 ... (<http://www.52im.net/forum.php?mod=collection&action=view&ctid=9>、<http://www.52im.net/forum.php?mod=collection&action=view&ctid=8>)

[2] 有关IM/推送的通信格式、协议的选择：

《为什么QQ用的是UDP协议而不是TCP协议？》<http://www.52im.net/thread-279-1-1.html>

《移动端即时通讯协议选择：UDP还是TCP？》<http://www.52im.net/thread-33-1-1.html>

《如何选择即时通讯应用的数据传输格式》<http://www.52im.net/thread-276-1-1.html>

《强烈建议将Protobuf作为你的即时通讯应用数据传输格式》<http://www.52im.net/thread-277-1-1.html>

《移动端IM开发需要面对的技术问题（含通信协议选择）》<http://www.52im.net/thread-133-1-1.html>

《简述移动端IM开发的那些坑：架构设计、通信协议和客户端》<http://www.52im.net/thread-289-1-1.html>

《理论联系实际：一套典型的IM通信协议设计详解》<http://www.52im.net/thread-283-1-1.html>

《58到家实时消息系统的协议设计等技术实践分享》<http://www.52im.net/thread-298-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=18>

[3] 有关IM/推送的心跳保活处理：

《Android进程保活详解：一篇文章解决你的所有疑问》<http://www.52im.net/thread-438-1-1.html>

《Android端消息推送总结：实现原理、心跳保活、遇到的问题等》<http://www.52im.net/thread-341-1-1.html>

《为何基于TCP协议的移动端IM仍然需要心跳保活机制？》<http://www.52im.net/thread-281-1-1.html>

《微信团队原创分享：Android版微信后台保活实战分享(进程保活篇)》<http://www.52im.net/thread-210-1-1.html>

《微信团队原创分享：Android版微信后台保活实战分享(网络保活篇)》<http://www.52im.net/thread-209-1-1.html>

《移动端IM实践：实现Android版微信的智能心跳机制》<http://www.52im.net/thread-120-1-1.html>

1.html

《移动端IM实践：WhatsApp、Line、微信的心跳策略分析》<http://www.52im.net/thread-121-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=17>

[4] 有关WEB端即时通讯开发：

《新手入门贴：史上最全Web端即时通讯技术原理详解》<http://www.52im.net/thread-338-1-1.html>

《Web端即时通讯技术盘点：短轮询、Comet、Websocket、SSE》<http://www.52im.net/thread-336-1-1.html>

《SSE技术详解：一种全新的HTML5服务器推送事件技术》<http://www.52im.net/thread-335-1-1.html>

《Comet技术详解：基于HTTP长连接的Web端实时通信技术》<http://www.52im.net/thread-334-1-1.html>

《WebSocket详解（一）：初步认识WebSocket技术》<http://www.52im.net/thread-331-1-1.html>

《socket.io实现消息推送的一点实践及思路》<http://www.52im.net/thread-188-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=15>

[5] 有关IM架构设计：

《浅谈IM系统的架构设计》<http://www.52im.net/thread-307-1-1.html>

《简述移动端IM开发的那些坑：架构设计、通信协议和客户端》<http://www.52im.net/thread-289-1-1.html>

《一套原创分布式即时通讯(IM)系统理论架构方案》<http://www.52im.net/thread-151-1-1.html>

《从零到卓越：京东客服即时通讯系统的技术架构演进历程》<http://www.52im.net/thread-152-1-1.html>

《蘑菇街即时通讯/IM服务器开发之架构选择》<http://www.52im.net/thread-31-1-1.html>

《腾讯QQ1.4亿在线用户的技术挑战和架构演进之路PPT》<http://www.52im.net/thread-158-1-1.html>

《微信技术总监谈架构：微信之道——大道至简(演讲全文)》<http://www.52im.net/thread-200-1-1.html>

《如何解读《微信技术总监谈架构：微信之道——大道至简》》<http://www.52im.net/thread-201-1-1.html>

《快速裂变：见证微信强大后台架构从0到1的演进历程（一）》<http://www.52im.net/thread-168-1-1.html>

《17年的实践：腾讯海量产品的技术方法论》<http://www.52im.net/thread-159-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=7>

[6] 有关IM安全的文章：

《即时通讯安全篇（一）：正确地理解和使用Android端加密算法》<http://www.52im.net/thread-216-1-1.html>

《即时通讯安全篇（二）：探讨组合加密算法在IM中的应用》<http://www.52im.net/thread-217-1-1.html>

《即时通讯安全篇（三）：常用加解密算法与通讯安全讲解》<http://www.52im.net/thread-219-1-1.html>

《即时通讯安全篇（四）：实例分析Android中密钥硬编码的风险》<http://www.52im.net/thread-312-1-1.html>

《传输层安全协议SSL/TLS的Java平台实现简介和Demo演示》<http://www.52im.net/thread-327-1-1.html>

《理论联系实际：一套典型的IM通信协议设计详解（含安全层设计）》<http://www.52im.net/thread-283-1-1.html>

《微信新一代通信安全解决方案：基于TLS1.3的MMTLS详解》<http://www.52im.net/thread-310-1-1.html>

《来自阿里OpenIM：打造安全可靠即时通讯服务的技术实践分享》<http://www.52im.net/thread-215-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=6>

[7] 有关实时音视频开发：

《即时通讯音视频开发（一）：视频编解码之理论概述》<http://www.52im.net/thread-228-1-1.html>

《即时通讯音视频开发（二）：视频编解码之数字视频介绍》<http://www.52im.net/thread-229-1-1.html>

《即时通讯音视频开发（三）：视频编解码之编码基础》 <http://www.52im.net/thread-232-1-1.html>

《即时通讯音视频开发（四）：视频编解码之预测技术介绍》 <http://www.52im.net/thread-235-1-1.html>

《即时通讯音视频开发（五）：认识主流视频编码技术H.264》 <http://www.52im.net/thread-237-1-1.html>

《即时通讯音视频开发（六）：如何开始音频编解码技术的学习》 <http://www.52im.net/thread-241-1-1.html>

《即时通讯音视频开发（七）：音频基础及编码原理入门》 <http://www.52im.net/thread-242-1-1.html>

《即时通讯音视频开发（八）：常见的实时语音通讯编码标准》 <http://www.52im.net/thread-243-1-1.html>

《即时通讯音视频开发（九）：实时语音通讯的回音及回音消除概述》 <http://www.52im.net/thread-247-1-1.html>

《即时通讯音视频开发（十）：实时语音通讯的回音消除技术详解》 <http://www.52im.net/thread-250-1-1.html>

《即时通讯音视频开发（十一）：实时语音通讯丢包补偿技术详解》 <http://www.52im.net/thread-251-1-1.html>

《即时通讯音视频开发（十二）：多人实时音视频聊天架构探讨》 <http://www.52im.net/thread-253-1-1.html>

《即时通讯音视频开发（十三）：实时视频编码H.264的特点与优势》 <http://www.52im.net/thread-266-1-1.html>

《即时通讯音视频开发（十四）：实时音视频数据传输协议介绍》 <http://www.52im.net/thread-267-1-1.html>

《即时通讯音视频开发（十五）：聊聊P2P与实时音视频的应用情况》 <http://www.52im.net/thread-269-1-1.html>

《即时通讯音视频开发（十六）：移动端实时音视频开发的几个建议》 <http://www.52im.net/thread-270-1-1.html>

《即时通讯音视频开发（十七）：视频编码H.264、V8的前世今生》 <http://www.52im.net/thread-274-1-1.html>

《简述开源实时音视频技术WebRTC的优缺点》 <http://www.52im.net/thread-225-1-1.html>

《良心分享：WebRTC 零基础开发者教程（中文）》 <http://www.52im.net/thread-265-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=4>、<http://www.52im.net/forum.php?mod=collection&action=view&ctid=5>

[8] IM开发综合文章：

《移动端IM开发需要面对的技术问题》 <http://www.52im.net/thread-133-1-1.html>

《开发IM是自己设计协议用字节流好还是字符流好？》 <http://www.52im.net/thread-150-1-1.html>

《请问有人知道语音留言聊天的主流实现方式吗？》 <http://www.52im.net/thread-175-1-1.html>

《IM系统中如何保证消息的可靠投递（即QoS机制）》 <http://www.52im.net/thread-294-1-1.html>

《谈谈移动端IM开发中登录请求的优化》 <http://www.52im.net/thread-282-1-1.html>

《完全自己开发的IM该如何设计“失败重试”机制？》 <http://www.52im.net/thread-280-1-1.html>

《微信对网络影响的技术试验及分析（论文全文）》 <http://www.52im.net/thread-195-1-1.html>

《即时通讯系统的原理、技术和应用（技术论文）》 <http://www.52im.net/thread-218-1-1.html>

《开源IM工程“蘑菇街TeamTalk”的现状：一场有始无终的开源秀》 <http://www.52im.net/thread-447-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=10>

[9] 开源移动端IM技术框架资料：

《即时通讯技术MobileIMSDK：快速入门》 <http://www.52im.net/thread-52-1-1.html>

《即时通讯技术MobileIMSDK：常见问题解答》 <http://www.52im.net/thread-60-1-1.html>

《即时通讯技术MobileIMSDK：压力测试报告》 <http://www.52im.net/thread-57-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=1>

[x] 有关推送技术的文章：

《iOS的推送服务APNs详解：设计思路、技术原理及缺陷等》 <http://www.52im.net/thread-345-1-1.html>

《Android端消息推送总结：实现原理、心跳保活、遇到的问题等》<http://www.52im.net/thread-341-1-1.html>

《扫盲贴：认识MQTT通信协议》<http://www.52im.net/thread-318-1-1.html>

《一个基于MQTT通信协议的完整Android推送Demo》<http://www.52im.net/thread-315-1-1.html>

《求教android消息推送：GCM、XMPP、MQTT三种方案的优劣》<http://www.52im.net/thread-314-1-1.html>

《移动端实时消息推送技术浅析》<http://www.52im.net/thread-288-1-1.html>

《扫盲贴：浅谈iOS和Android后台实时消息推送的原理和区别》<http://www.52im.net/thread-286-1-1.html>

《绝对干货：基于Netty实现海量接入的推送服务技术要点》<http://www.52im.net/thread-166-1-1.html>

《移动端IM实践：谷歌消息推送服务(GCM)研究（来自微信）》<http://www.52im.net/thread-122-1-1.html>

《为何微信、QQ这样的IM工具不使用GCM服务推送消息？》<http://www.52im.net/thread-117-1-1.html>

>> 更多同类文章 ... <http://www.52im.net/forum.php?mod=collection&action=view&ctid=11>

[y] 更多即时通讯技术好文分类：
<http://www.52im.net/forum.php?mod=collection&op=all>

作者：Jack Jiang (点击作者姓名进入Github)
出处：<http://www.52im.net/space-uid-1.html>
交流：欢迎加入即时通讯开发交流群 215891622
讨论：<http://www.52im.net/>
Jack Jiang同时是【原创Java Swing外观工程BeautyEye】和【轻量级移动端即时通讯框架MobileIMSDK】的作者，可前往下载交流。

1人 喜欢 喜欢

回应 推荐 喜欢 只看楼主

你的回应

回应请先 登录, 或 注册

☐ 推荐到广播

