

腾讯“一秒钟法则”的最佳实践

原创 2017-02-16 刘昕 高效运维

点击蓝字，轻松关注



作者简介：



刘昕

腾讯 无线运营部副总经理

负责QQ浏览器、应用宝、腾讯手机管家、腾讯电脑管家的运营规划、平台研发工作。

对移动互联网应用的用户感知与技术指标的度量、优化有丰富的实践经验，专注于业务架构优化、弹性伸缩、运营服务管理、运营体系建设，帮助产品打造极致的技术基础和质量口碑。

1、前言

首先做一下自我介绍，我叫刘昕，来自腾讯 MIG 的无线运营部，我这个部门会支撑腾讯 MIG 的业务系统，这个业务系统给用户提供的产品包括应用宝、手机浏览器、手机管家、电脑管家、腾讯地图这样的一些产品以及服务，基本上是工具类产品，我之前也负责过手机 QQ、手机游戏的运维和优化工作。



GOPS 2016 全球运维大会·上海站

GOPS 2016
Shanghai

在这里我想给大家分享一下小故事：

话说三年前，腾讯第一个在微信和手机 QQ 上的移动游戏，我不知道大家是否还记得那个游戏，天天爱消除，第一天在微信的平台上上 IOS 客户端，在手机 QQ 上进行安卓客户端发布。

第一天的运营效果，两个产品平台推广用户，结果就出问题了，问题是在微信上的 IOS 客户端用户量非常大，在手机 QQ 上安卓的客户端游戏用户量特别小，可能差一个数量级，一个是新增几百万，一个新增几十万用户。

这个时候 QQ 平台的老板就坐不住了，难道真是 QQ 平台不好吗？即便不好，但是安卓客户基数很大，就反过来质问说，QQ 的新用户下载转换率太低了，是不是我们 CDN 的服务不好？

我和他说，其实服务是很好的，因为我在应用宝里面用你的服务做应用下载，其实下载成功率、用户转化率是能够看清楚的，要比你在这次游戏推广的过程当中要高很多。

为什么呢？我们做了分析，这中间出现的问题有很多，用户的下载成功率可能和终端的类型有关，也可能和用户连接的网络有关，也可能和 CDN 服务有关，也可能和 CDN 服务下载要素有关。

我们发现这个过程当中有比较多的问题，在下载成功率上有几个百分点的差异，怎么解决这个问题？因为已经上线了，第二天又是一个很大的推广，怎么办？

当时只能找临时的解决方案，临时修改一下 CDN 调度逻辑，从原来的域名就近访问，变成用户来源 IP 的就近访问，做一些简单的优化，其实本身客户端的下载逻辑和服务逻辑也都是有一些问题的，但是已经改不了的。

我们简单做了这样一个调整，第二天第三天的效果比较好，运营数据也逐渐跟上来了，我讲的这个故事想表达的意思就是说我们在移动互联网给用户提供服务的。


关于用户体验以及应用的性能管理，里面有很多是和终端网络云端服务相关的，所以说我们再衡量和评估整个用户体验我们需要从端到端进行管理，我们在各个环节都需要有很多的基础解决方案和管理方式。

在这个过程中想运营的产品，有些人说靠设计靠产品运营靠产品，我们作为技术的运营部门或者是支撑部门。

我们希望通过给用户极致的体验才是一个产品真正达到精品的基石，产品可以有一个好的代言。

但是如果你的产品体验非常糟的话，用户很快就会走，你能吸引到最新的用户，但是你满足不了用户的基本需求，你的体验达不到竞争对手，或者说不能超越竞争对手，用户很快就会走。

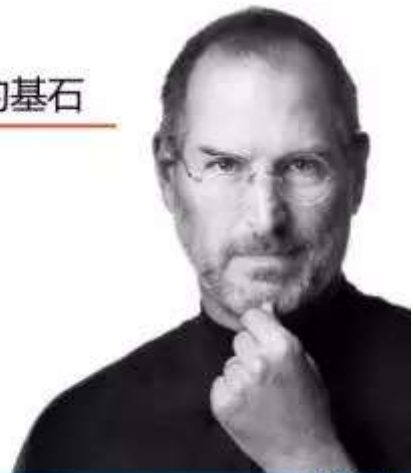
特别是对于我所复杂的这几个腾讯的工具类产品，基础的用户体验都是很重要的，比如说下载的成功率、下载的速度、网页打开的速度以及成功率，这些都会最终影响用户是否还留在你这儿，还在使用你的产品。

 精品是设计驱动

 精品靠产品运营

 精品是情怀

给用户极致的基础体验才是精品的基石



GOPS 2016 全球运维大会·上海站

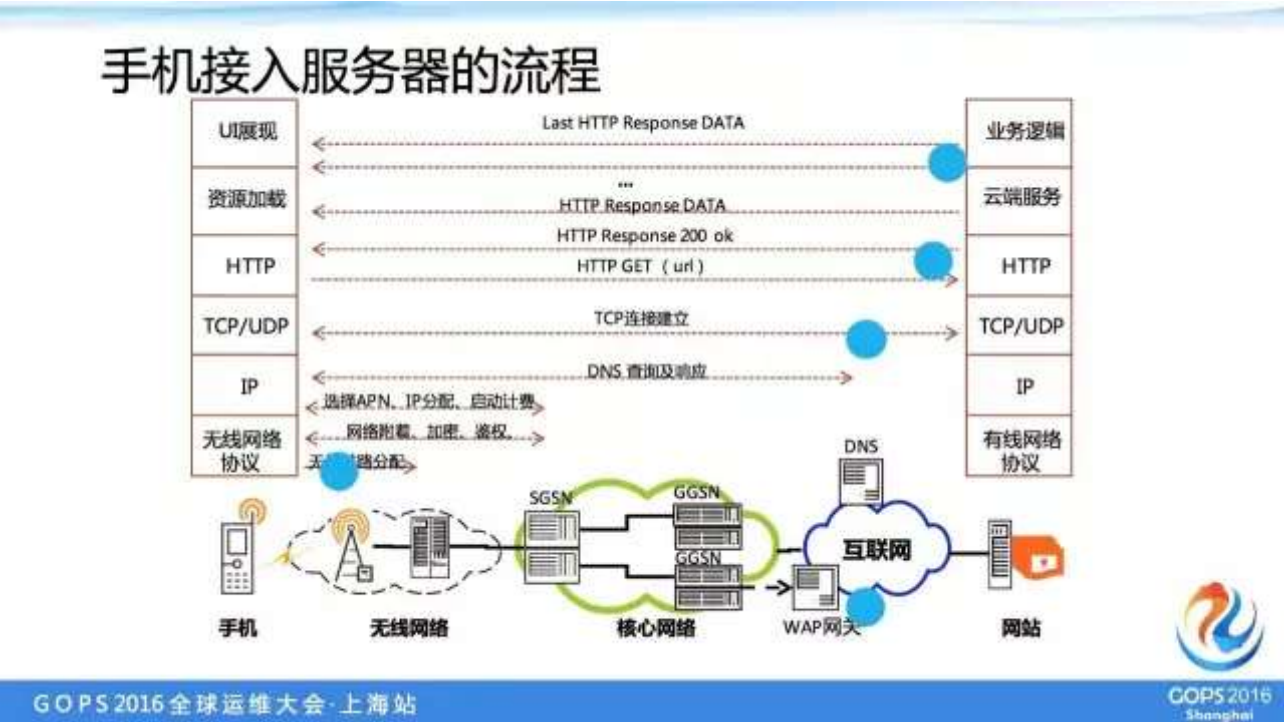
GOPS 2016
Shanghai

本文大概分三部分：

- 介绍一下什么叫一秒钟法则，我会简单介绍一下移动网络的特点和特性，以及一秒钟法则。
- 第二个部分告诉大家怎么做，怎么在移动互联网环境里面，真正给用户做到好的体验。

- 最后讲一下我在运营产品的时候到底是什么样的管理方式，能够持续的不断优化。

2、移动互联网特点和 “一秒钟法则”



首先这页图给大家看的是一个手机的应用访问后台服务经过的网络单元，你的网络承载情况是什么。我们看到这里面从手机到基站到手机基站控制器到运营商的核心网络，到互联网，最终才能到达云端服务，我给大家简单介绍一下这个过程的流程。

比如说你要打开一个 app，要进行上网的时候，打开 app 和我们传统的互联网有线互联网服务不一样，有线的互联网服务只要连上网络，网络就是通的，输入域名就可以把你的请求发送到后端。

但是在无线环境里面是不一样的，手机客户端如果要发送一个网络请求之前必须要和无线网络连上，要在基站连到对应的基站，同时申请一些无线电路实现数据的传输，如果没有无线电路分配，虽然连在网络上，你是没有办法再进行数据传输的。

连上以后还有一些其他的附加过程，比如说认证，移动的手机是不是连在移动的基站上，手机有没有欠费之类的工作。之后才能真正完成基本的联网工作。接下来才是终端 IP 的分配，然后启动你的流量计费。

完成这样的一个工作，才能完成基本的网络接入，才能启动通常意义上有线互联网的网络请求。

所以说在移动网络，移动互联网里面，对于网络的连接和处理和有线是不一样的，需要更复杂一点，在各个环节其实都有相应的技术的门槛或者是其他的制约。

比如说在无线网络的无线电路分配上面，在移动互联网的基础 DNS 承载上面，在后台服务的连接上面，包括外部页面的处理和转化上面，都有和有线互联网不一样的地方。

2.1 移动互联网的特点

简单的我们总结几条，移动互联网基本特点，简单总结几条。

2.1.1 无形的接口提供物理连接

这句话怎么理解，我们搭建一个互联网服务的时候，比如说我们在后台服务上面看到一个用户连接上了和后台服务建立连接，我就认为这个连接一定是存在的。

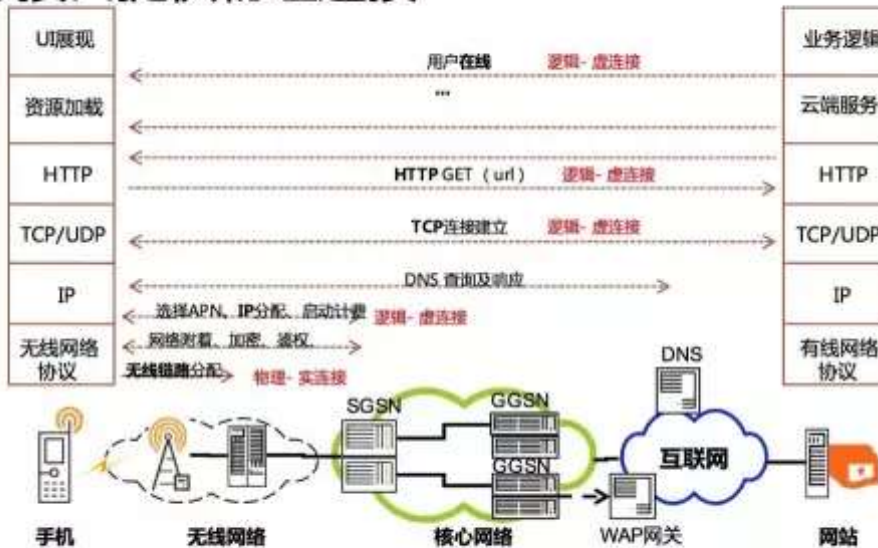
我有数据交互的时候从服务器后台发放数据，或者从客户端发送数据的时候，就基本上会认为我的数据对方是可以接收到的。

对于网络质量还有一定的基础，基本上一定可以接收到，但是无线网络是不一样的，所以在上层的 IP 连接、HTTP 连接，这些连接都是虚拟的。

你看到的用户在线状态不是确定的，都不是确定的，你只有在用户的终端手机和获得一段实际的无线电路资源分配的时候，你真正的建立和端到端的连接。

所以说在有线网络在物理场的连接才是实际连接，在上层的连接都是虚拟的连接。无线的接口提供实际的物理连接。

无形接口提供物理连接



GOPS 2016 全球运维大会-上海站

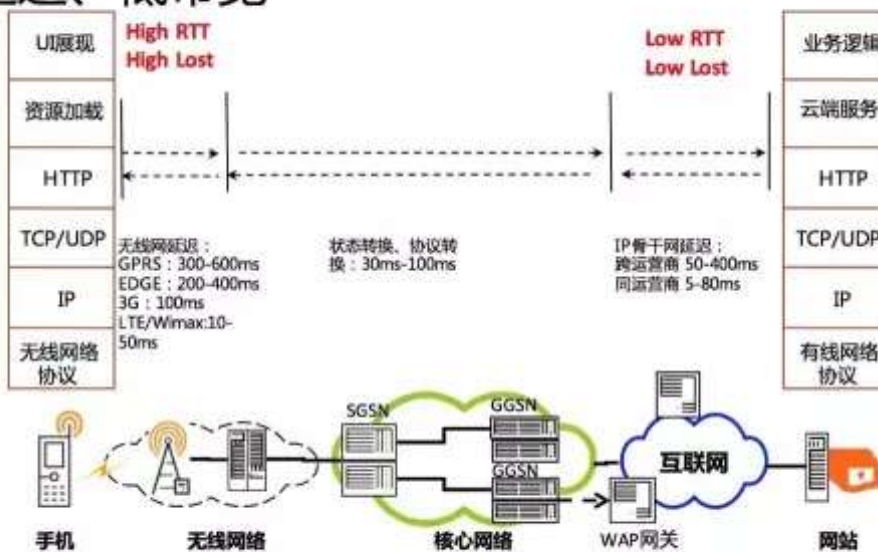
GOPS 2016
Shanghai

2.1.2 在移动互联网里面，网络特性是高延迟、低带宽

在手机和在无线网络连接过程当中，其实在所谓的延迟消耗上更多的发生在无线网络这一侧，当然无线网络为什么会出现这种情况，有几个问题。

首先是无线信号的传输有多径传输的问题，无线数据的发送需要把一定单位时间内的数据组合成一个大的数据或者说大的数据包，然后再进行传输，以及还有一些物理层的信号纠错、重传。

高延迟、低带宽



GOPS 2016 全球运维大会-上海站

GOPS 2016
Shanghai

另外就是确实你的传输带宽是比较低的，本身数据延迟就是存在的。

我们大概总结了几个特点:

1. 在 2G、2.5G 的情况下，大概是两百到六百毫秒的延迟
2. 3G 做到六百毫秒
3. 在 4G 情况下，大概是10毫秒、20毫秒之间
4. 未来 5G 可能希望做到一毫秒之内

我们可以简单的比较一下，比如说在家里固定宽带上网，我的电脑和运营商机房接入的设备基本上是在一到三毫秒之间的延迟。

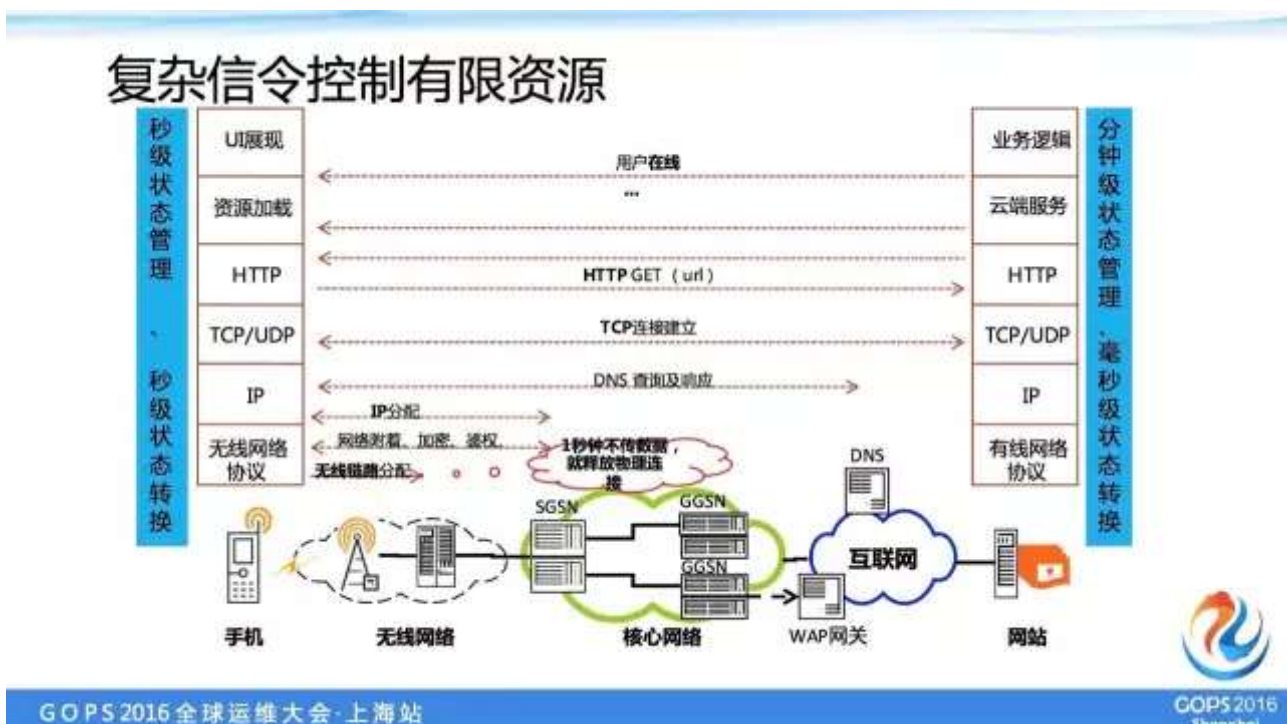
另外在无线网络里面各种状态的转换和协议的转换也需要消耗时间，剩下在IP骨干网上的传输是一样的，同运营商就是几毫秒之间，跨运营商就是几毫秒到几百毫秒之间。

2.1.3 复杂信令控制有限资源

无线网络的带宽资源都是有限的，最终决定就是带宽以及你的调制技术决定了你各种 2G、3G、4G 传输的最大接入带宽，而且这个接入带宽在一定范围内是很多用户共享的，并不是独占的。

比如说我们在一个小区内或者一个基站覆盖的所有用户，其实我们共享这个基站的传输带宽，通常意义上 4G 可以达到一百兆到两百兆，指的是一个基站下面一个用户独占资源可以达到这样的下行带宽，如果有很多用户，其实大家是共享这个带宽的。

同时我们在无线电路有分配也有释放的过程，其实在这个过程当中，这些网络设备用一套复杂的信令机制进行信道资源的分配和释放来进行控制的，包括手机的基站信号强度、链路的分配和释放。



所以在这样的环境里面，资源是有限的，资源的分配和释放必须要有一定的规则，我们大概总结的是在无线网络的环境里面，如果你获得分配的资源超过一定的时间，就按秒来计算，如果没有再进行数据交互的时候，你当前获得的无线电路的资源会被回收和释放的。

比如说我已经和网络建立了一个连接，有数据包的发送，如果连接没有持续的进行，我连接的对话保持，但是没有实际的数据包交互的时候，已经分配的信号资源无线电路的资源就会很快的释放掉，但是不同制式的网络，对于电路资源的管理方式是不一样的。

我们大概总结了一下，就是在移动互联网里面，状态的转换，比如说获得信号资源、线路资源和释放线路资源大概是两级的状态，而且状态之间的转化也需要秒级的时间来完成。

2.2 “一秒钟法则”

对于传统的有线互联网来说，我的状态的管理是按分钟计的，举个例子，一个 TCP 连接，连接网络时间通常默认是三分钟，有的可能调到十几分钟做一些优化。状态之间的转化是毫秒级的，比如说建立一个连接或者说连接终端释放，这些都是毫秒级的状态转换。

所有这些就会导致在移动互联网里面，所有的用户体验很难找到很具体的，很难用原来的方法进行保证，所有的技术手段和管理方式需要进行重新的定制和选择，我们在这个基础上，自己定义了这样一个原则，叫做**一秒钟法则**。

我们把一个页面的打开流程做了下面的分解：DNS 查询、建立连接、后台响应，页面加载，到客户端的解析排版再到 UI 展现，客户端出现首字首屏。

我们认为什么样的情况下，在移动互联网里面给用户最终的体验是合适的，我们所有的用户的体验管理按照什么样的标准来巩固呢？就是一秒钟法则。

简单地来讲，在不同的网络环境里面是不一样的：

- 说在 2G 网络里面，我们希望做到 DNS 解析和客户端的连接建立时间控制在一秒钟之内；
- 3G 的网络如果是浏览页面打开，我们希望是首字时间在一秒钟之内；
- 如果是 Wifi 或者是 4G 我们希望首屏的时间刷新一个页面的时间在一秒钟之内。这是我们总结出来的一些原则。



3、怎么做才给用户好的体验

接下来我给大家探讨一下我们实际在运营过程当中我们通过什么样的方式去优化体验，实践我们所谓的一秒钟法则。

分为三个部分：

- 接入调度优化
- 协议的优化
- 业务逻辑的优化

3.1 互联网接入优化

3.1.1 移动互联网接入调度尽量减少 DNS 影响

简单给大家列举一下情况，比如说 DNS 的承载是不一样的，在有线的互联网里面，我知道这个 DNS 服务器，大概知道所能够提供解析的用户所在区域，在移动网络里面基本上不是这样的，很多 DNS 承载全国全网的用户。

另外还有一些 DNS 设置的问题，还有一些 DNS 的劫持以及解析成功率低的问题。

第二个要务就是说不要把接入调度的很多判断放在客户端，比如说客户端我们能获得的信息的可信度，简单地来说，比如说我们这个客户端获得当前网络连接的状态，是 Wifi 的还是 2G 的、3G 的，可信度并不高。

同时，在我们很多客户端发布了以后，任何小的修改，我们在客户端需要更新来完成的时候，在手机客户端版本更新的周期非常长，需要很多其实的方式来解决版本更新的问题。

第三，在客户端里面其实有很多写死IP的问题，还有一些本身手机客户端并不是一个真正合理的用户，中间会有很多黑终端、山寨的终端、山寨机这样的一些情况。

简单介绍一下在接入调度的两条路线，以及在这两条路线的演进过程。

移动客户端接入调度有两种主流：

- 传统的 DNS 解析
- 在客户端配置你的服务器列表

在移动互联网里面对于一个服务的部署，可能需要更多的接入点，甚至说更多的域名，因为很多域名经常会被污染掉或者是错误解析，如果希望你的用户不受影响，你需要更多的域名，用域名的方式调动更新。

3.1.2 设IP列表，需要动态的方式更新服务器列表

而不是写死了以后放在那里，逐渐有一些方法的融合，比如说用一些域名解析，用一个域名获得这个 IP，然后再去直连对应的服务器 IP。还有一个就是对于手机客户端的协议，你一个应用最好能够有多端口多协议的组合，不能把所有的宝压在一个协议。

比如说你是一个长连接的协议，可能对应的端口号连接类型可能需要一个动态的变化的过程，而不是完全用一种方式，因为在有些网络环境下，可以用 TCP 连接的。

有些网络环境下只能用 IP 连接之类的，各种网络接入环境是不一样的，也是比较复杂的，需要有一些手段来保证用户在任何接入环境下都能够使用我们的服务。

3.1.3 终端测速提供最快接入点

还有比如建立一些网络，不需要每次用户连接到网络上都需要重新解析IP重新获得服务器列表，举个简单的例子，我们的手机在连接网络的时候，基本上每天平均大概会有20多次的网络切换。

比如说连接到 4G，连接到 Wifi，连接到这个 Wifi 或者那个 Wifi，有这么多网络切换的时候，不需要每次都重新测速选择一个最新的列表，我可以做一个简单的，用 4G 的时候连到哪个服务器是最快的。

连到公司 Wifi 连在哪个接入点是最快的，在家里的 Wifi 哪个接入点是最快的，这样就可以建立和后台建立连接。这是我们的一些实践过程，我们通过一些测速的数据分析以及处理。

然后找到最优的接入调度解决方案，我们可以把颗粒度做到国内省份运营商选择我们最优的接入路由，从就近接入到就快接入，我们不太关心用户和后台服务是不是同一个供应商，我们只需要选择最快的接入点就可以了。

移动应用接入调度优化：就快接入



GOPS 2016 全球运维大会·上海站

GOPS 2016
Shanghai

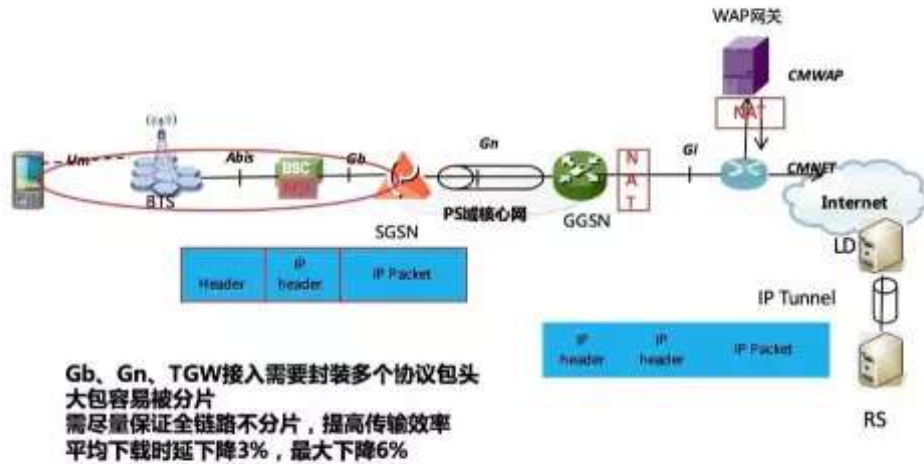
这个的前提就是你需要一套精准的 IP 库，IP 库的运营是一个长期持续的过程，回到刚才讲的案例，我们在解决游戏的下载调度的时候怎么解决，我们把传统的 DNS 解析找到最近的 CDN 下载点这种方式，改成了根据用户的终端 IP 判断最近的 CDN 业务，这样的话下载质量下载成功率就可以提高很多。

我们现在维护的 IP 库对国外运营商的准确率 98% 以上，国内城市准确率在 97% 以上，其他国家、省份、运营商、网络接入类型都超过 99%。

3.2 协议优化

协议优化这里面简单介绍一下一些协议优化原则，比如说要快速关闭 TCP 快速回收，初始 RTO 时间尽量不低于十秒这样的一些原则。

关掉TCP快速回收：避免分片、乱序



GOPS 2016 全球运维大会·上海站

GOPS 2016
Shanghai

这里面会简单介绍一下为什么有这样的原则，其实这里面的各种技术的技巧有很多，为什么关掉 TCP 快速回收，因为网络环境里面有多重分包和解析，传输数据包过大的时候有很多包会重新打包，同时传输的时候数据包会发生变化和不一致，我们用 TCP 快速回收做连接管理的时候就不会出现这个问题，如果提供手机服务就把这个 TCP 快速回收关掉。

初始 RTO 时间也是一样的，在手机环境下，建议设成一秒钟，就是谷歌前几年发了一个文章，给大家一个建议，结果我在我们实际调研过程当中发现，在弱网络的环境下，如果 RTO 时间是一秒的时候 TCP 连接率非常低，数据传输可靠性比较低，我们建议设到三秒以上。

初始拥塞窗口，小页面的环境下，把拥塞窗口调大一点，就不会受到 TCP 慢启动的影响，可以在很快的时间完成小页面数据传输，页面大小在1M以下的时候这个原则基本上是适用的。

协议优化——更快

关闭TCP快速回收

Init RTO时间不低于3秒

初始拥塞控制窗口不小于10

Socket buffer增大到64K以上

TCP滑动窗口可变

控制发包大小1400字节以下，避免分片

协议优化选择更高效率的协议，大家只需要这几个原则就好了，具体的协议到每个公司每个行业都不一样，第一我们希望能够做连接的重用，控制并发连接，超时控制，包头精简，内容压缩，选择越高效的越好。

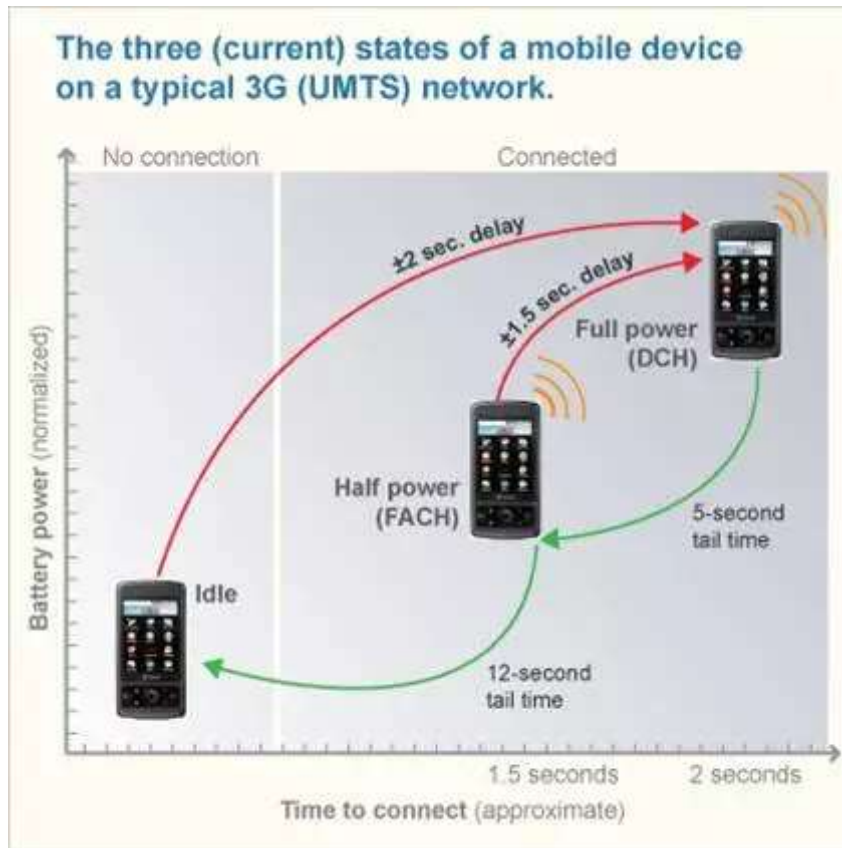


3.3 业务逻辑优化

在长连接的如果做到又省流量又省包量又省电量，需要控制一下长连接的心跳时间在一个合理范围，手机的终端耗电主要是三个部分，一个是屏显一个是内存第三个就是网络连接。

随着大屏手机出现，对于屏显耗电部分会越来越多，基本上在之前的就是三分天下，在不同的网络制式下，手机的发射功率是不一样的，手机耗电在通讯的时候耗电取决于连接什么网络，连接到 Wifi 网络相对省电，连接 3G、4G 会更耗电，以及和数据后台的交互次数和频率，还有你的状态转换，最后还有信号强弱，网络信号的强弱。

下一张图介绍一下耗电量的示意图，我们看到中间突出来的台阶部分，就是手机发送数据包的时候的耗电量。数据包不发放的时候经过几秒钟的时间就下降到闲置的状态，耗电量是下降的。



通过这样的一个对比，这是 3G 网络里面的手机耗电量以及手机状态的示意图，纵坐标就是耗电量大小，当我的手机工作的时候耗电量最大，如果不工作的时候切换到其他状态，这个时候手机的耗电量是非常小的。

切换的时间，当我的手机工作的时候，经过五秒钟以后，手机就会设置到 half power，再经过十二秒就会再设置到 Idle 状态，这时把之前所有分配的无线链路的资源都释放掉。这是耗电量的对比，在超过三百秒的心跳间隔时间是一个比较稳定的低电量消耗的状态。

还有一些业务逻辑优化，这里面简单介绍一些原则，比如说简化逻辑，消息的操作交互的次数频率尽量减少，时间间隔超过一秒钟的这种通讯业务逻辑一般建议合并，不是过几秒钟又和手机后台进行一次交互。

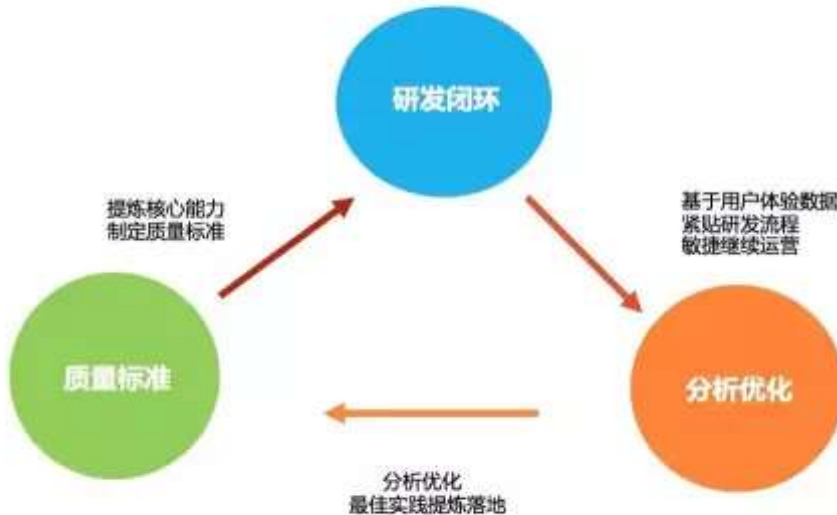
其他的一些逻辑在不同的特定网络环境下都有不同的设定技巧，比如说下载环境的处理如何获得资源包大小，有一些特殊的技术的处理方案。

前面给大家介绍一下技术优化手段，这些技术优化手段具体的体现，在我同时张丹关于手游的优化里面会讲的更详细一些，对于 DNS 的劫持处理怎么实现，怎么保证用户最好的体验。我这里主要讲基本原则。

4、对用户体验的优化

关于用户体验优化的方法，其实主要是三个步骤，首先我们必须建立一个质量定义标准，上午我们也看到了谷歌的那个也会讲 SLR、SLO，我们会定义一些和用户体验相关的标准，用户打开一个页面的时间，我们定义的标准是和用户直接的感知和体验相关的，而且这种指标是可以衡量到手机终端到后台的端到端的质量的，而不是简单的衡量后台的质量或者终端的质量。

用户体验提升实践



在研发的刚才当中我们不断加入这些质量标准，同时不断强化质量的要求，在实际运营过程当中我们也会不断收集在用户身上实际表现是什么样，因为是端到端的用户体验。最终我们会总结一些实际可以操作的能够落地的优化方案来进行。

我们简单介绍一下几个关键路径，第一个就是说我在每个产品的核心用户体验里面我会进行用户的获取、激活、留存以及收益这几个阶段定义这些关键阶段的一些关键的指标，以及这些阶段转换的关键指标。

举一个简单的例子，应用宝的应用分发下载，我们这里在用户获取展示的时候，我们会衡量应用宝的页面打开时间，打开的成功率，用户的激活下载我们会衡量用户的下载成功率，衡量用户下载的完成率，下载完了以后我们会衡量下载 app 软件在用户手机端的安装成功率，之后我们会衡量用户的激活率。

你的应用下载完成之后有没有转化成对应的活跃用户，我们建立这样一套完整的数据体系，所以说比如说我们在应用宝这里面应用页面加载一秒钟法则，在服务里面会考虑连接的触达率，到后台的用户消息是不是能够被用户真正收到。

在实际的收益这个阶段我们会考虑当有商业收入的时候，比如说你的登录支付是否成功，针对于负责的不同的几个产品不同的几个阶段对应一些核心的指标，而且这些核心指标是可以通用的，在不同的产品之间是可以通用的。

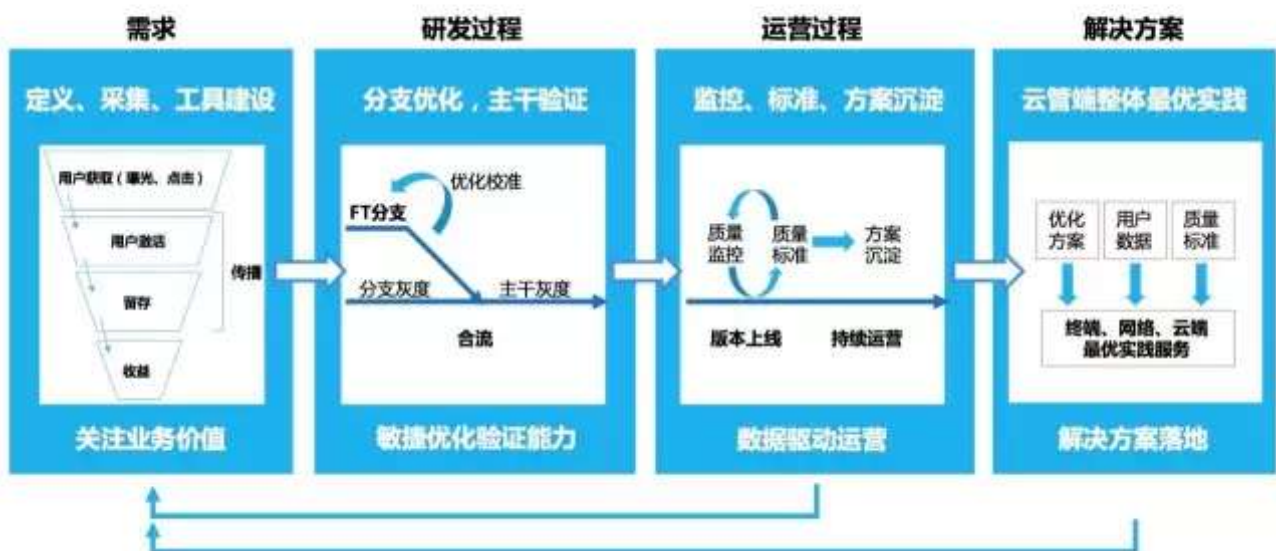
关键路径一：提炼核心能力，设定质量标准



第一个路径就是说前面我们起了个项目“魔镜”，去度量我们最终的用户体验，第二个过程我们和研发流程结合，在核心的业务指标里面，我们在研发比如说分支的开发、合流的过程当中我们定义一些质量标准。

如果达不到不允许进行发布，发布了以后我们也会通过一些方式验证当前的版本或者说之前做的一些优化手段有没有得到改善。最终在用户发布的时候我们会采集客户端的用户指标信息，然后衡量我的运营标准和客户体验。

关键路径二：敏捷研发、持续运营、数据驱动



在这里面我们起了一个名字“灯塔项目”，完成客户端业务指标的打点上报后台数据收集以及处理的工作，很多的业务指标基本上可以实现实时看到对应的结果。在测试过程当中其实也是一样

的，同样的方案在不同的人群里面做对应的灰度，观察一个方案落实的时候，在前后有没有产生优化的效果。

第三个就是提炼的最佳解决方案，对于下载服务、接入调度服务、长连接服务，我们提供了手机终端 SDK 的集成和后台服务的配合，当一个应用完成一个下载功能的时候不需要自己开发，只需要把下载的 SDK 集成进去，就可以保证有很好的下载体验和质量。

关键路径三：提炼最佳实践落地、哈雷项目



这就是一些优化的效果以及改进的过程，比如说对下载问题的解决，我们分析了下载过程当中有用户取消也会有下载劫持，下载失败当然有很多原因，终端后台的原因还有一些启动原因。

我们每个版本都会尝试解决一部分失败原因，最终让下载完成率提升14%，下载成功率提升2个百分点，最终下载成功率达到的99%，下载完成率达到96%这样的程度。

这个也是我们对于手机浏览器的一个体验和页面访问速度提升的一些解决方案，在不同的环节应用不同的手段来解决问题。

这是实际效果，比如说我们在手机客户端，这是几个不同的版本，5.1、6.1、6.9，左边是耗时的平均时间，上午大家听到了谷歌的工程师也不完全是考虑平均时间，其实也看百分比，我们定义了一个80%的范围，80%的请求都在一秒钟以下，每个版本的比例，当平均值达到的时候，更多的是提高那个百分点。

这是一个启动耗时优化，这是我们在实际运营刚才当中发现比如说应用宝的 app 启动，大于一秒钟的时候，用户其实就会是流失，用户的耐心是有限的，这也是我们为什么提出一秒钟法则。

app 启动耗时三秒钟的时候就会有一个阶梯性上升，从空白页面到加一个默认页面，会挽回一部分流失用户，即便我们做不到一秒钟的时候，这是我们在实际过程当中的一些优化手段。

近日好文：

《Hadoop 3中的磁盘管理大招解密》

《携程网基于应用的自动化容量管理与评估》

《日请求从百万到八亿的技术历程》

《Gitlab 删库引发的思考：你也陷入到“人肉运维”的怪圈中了？》

GOPS2017 · 深圳站，鹅厂有很多技术要揭秘



GOPS2017·深圳站

**这次
深圳的大厂按捺不住了
都想拿出新技术来炫耀一下**

- 会议地点：南山区圣淘沙酒店（翡翠店）
- 会议时间：2017年4月21日-22日

现在购票，您将享受 **88折（截止2月19日）**，您可点击“[阅读原文](#)”链接直接享用：

阅读原文
