对网易云音乐参数（params，encSecKey）的分析

我们如果对网易云音乐进行爬虫的话，我们会发现，提交的参数是（params，encSecKey），然而这两个参数是一串很长的东西



我们要对网易云进行爬虫，那么就一定要将这两个参数弄明白，然后才可以进行爬虫，不然这两个参数将限制着我们很多事情，下面是我对这两个参数的分析，参考了知乎大神的方法，这里将原帖地址写上来https://www.zhihu.com/question/36081767

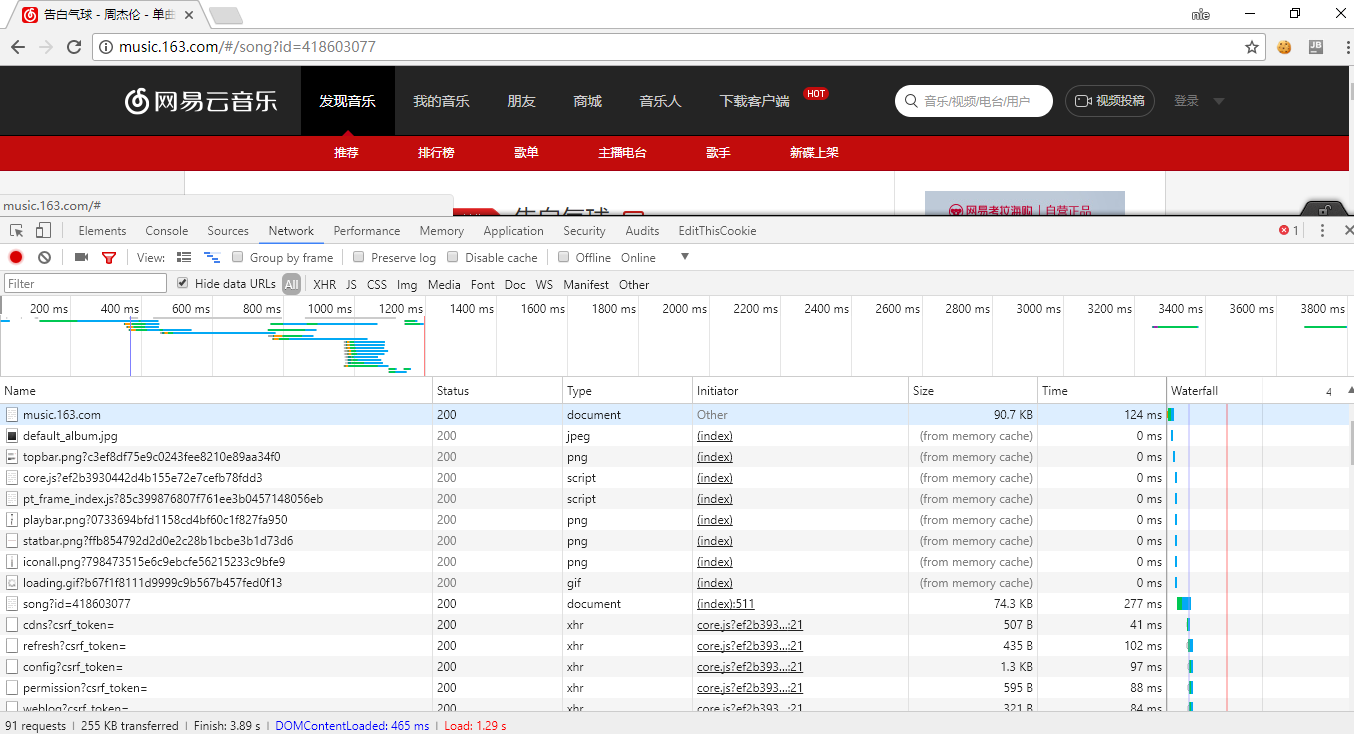
主要参考 作者：[平胸小仙女](https://www.zhihu.com/people/luocaodan" \t "https://www.zhihu.com/question/_blank) 的分析

下面进行分析：

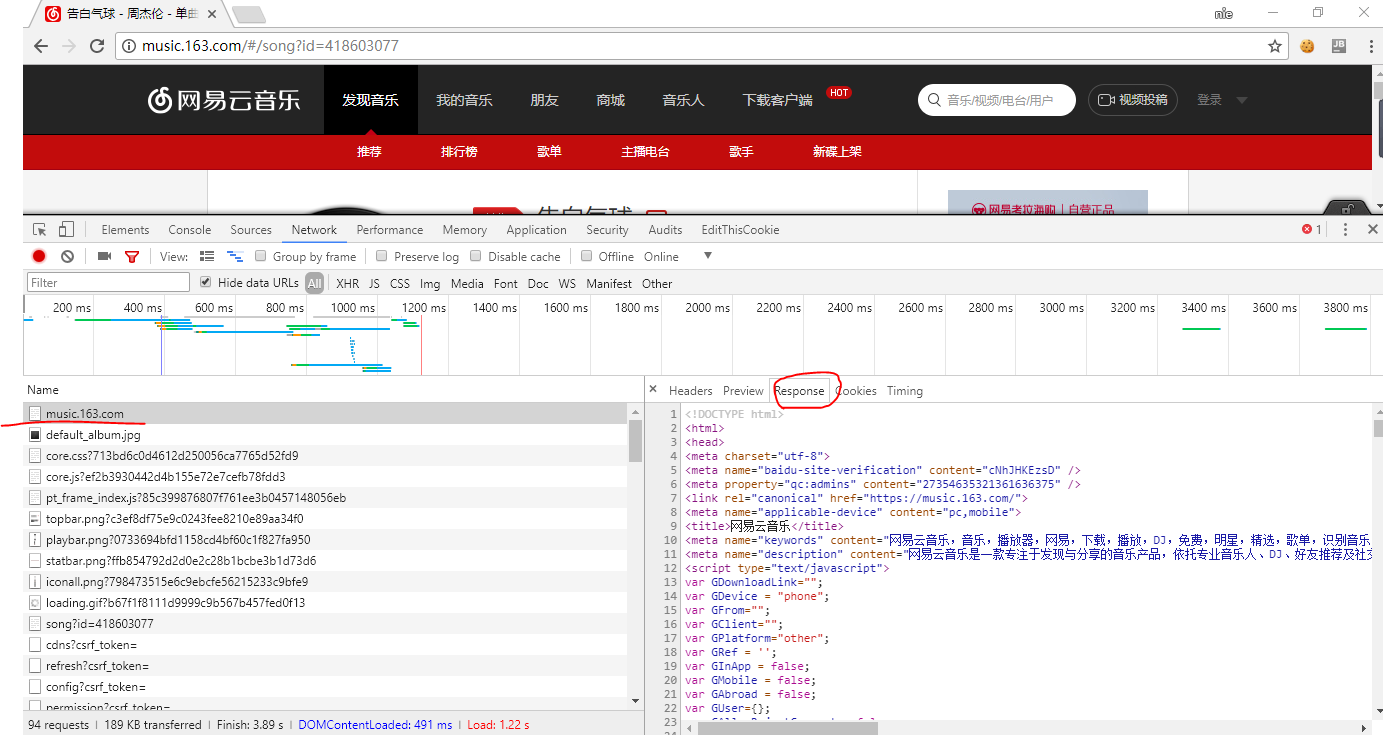
我在分析中要用到的几个软件，Google浏览器，火狐浏览器，burpsuite，Fiddler4，sublmie text3，notepad，pycharm

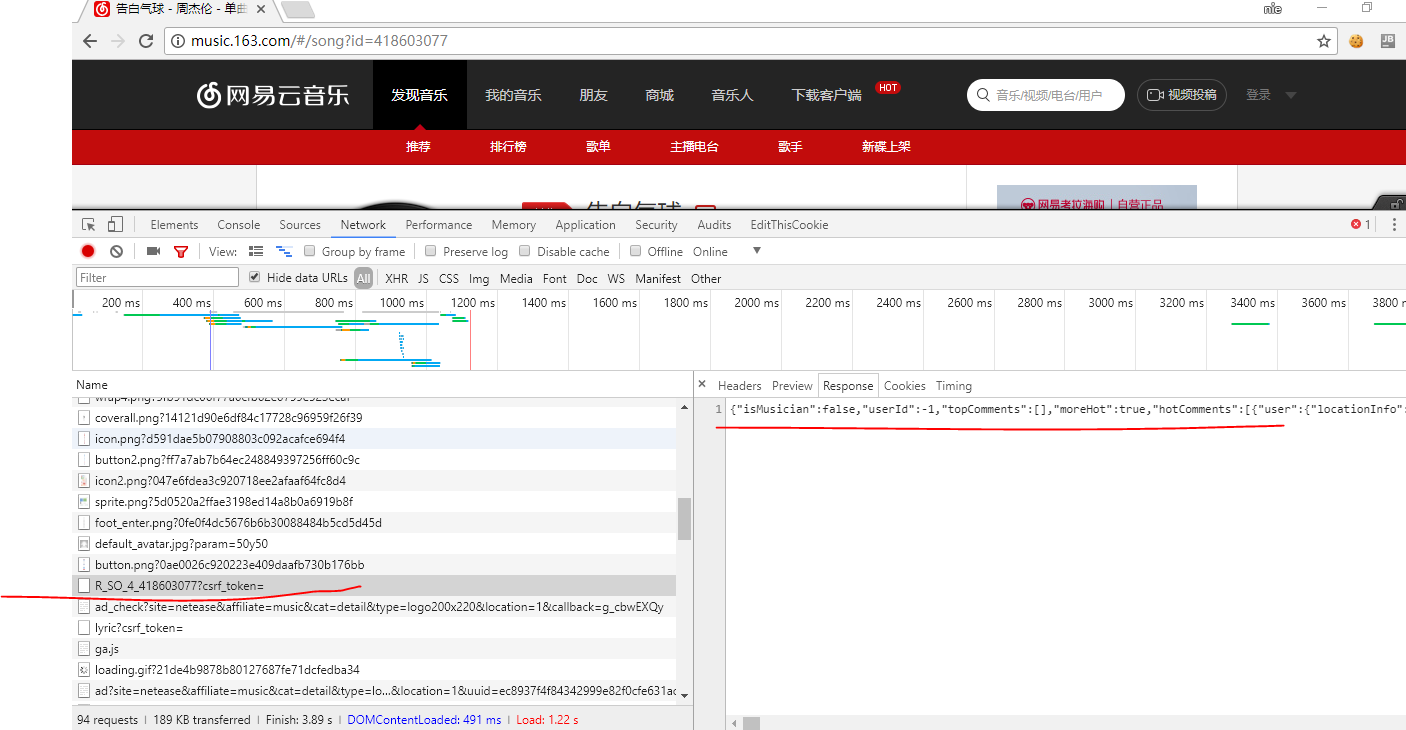
例子是：周杰伦的告白气球，爬取评论

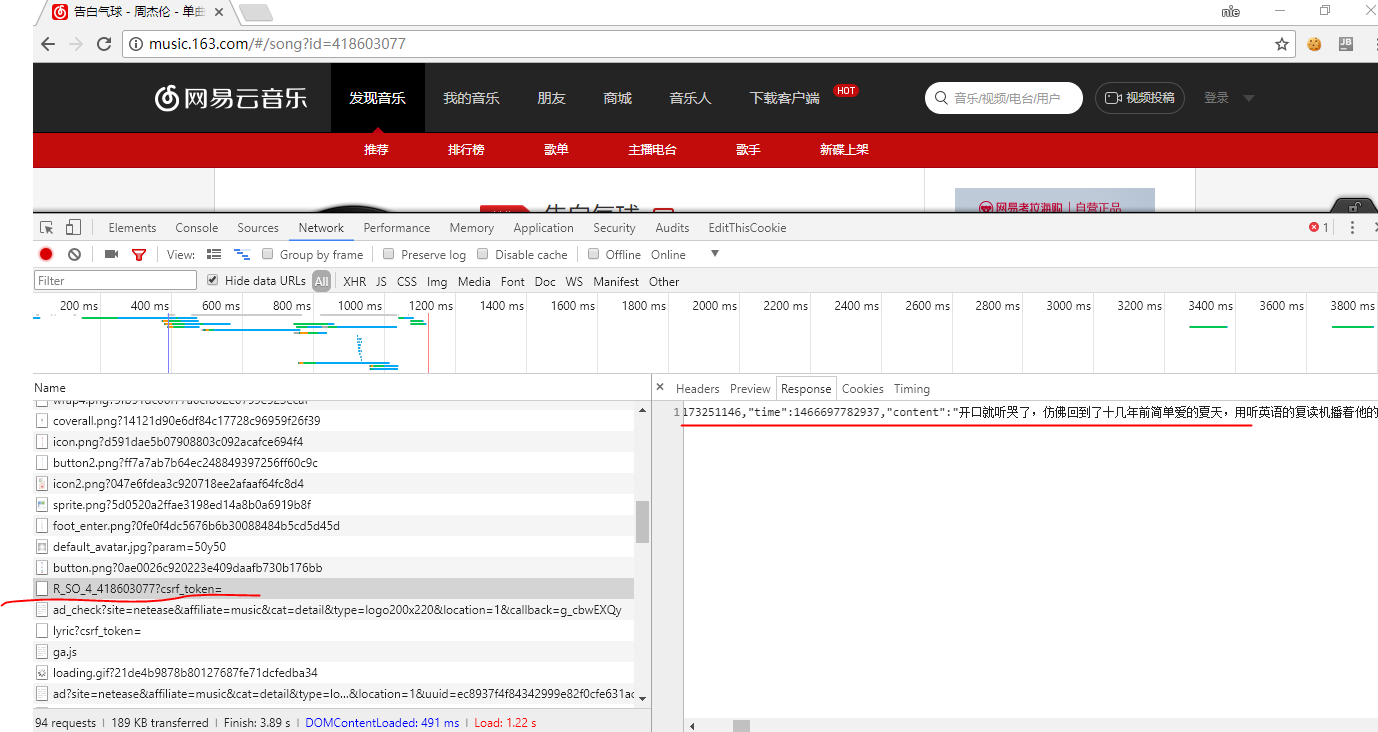
首先我们用Google浏览器查看（F12）



点击第一个name，然后选择response

看里面的内容，然后一直再name里面找，知道找到一个在response里面评论的地址





这里可以看到

Name为R\_SO\_4\_418603077?csrf\_token=

Response里面为评论

查看headers知道url为

<http://music.163.com/weapi/v1/resource/comments/R_SO_4_418603077?csrf_token=>

参数为

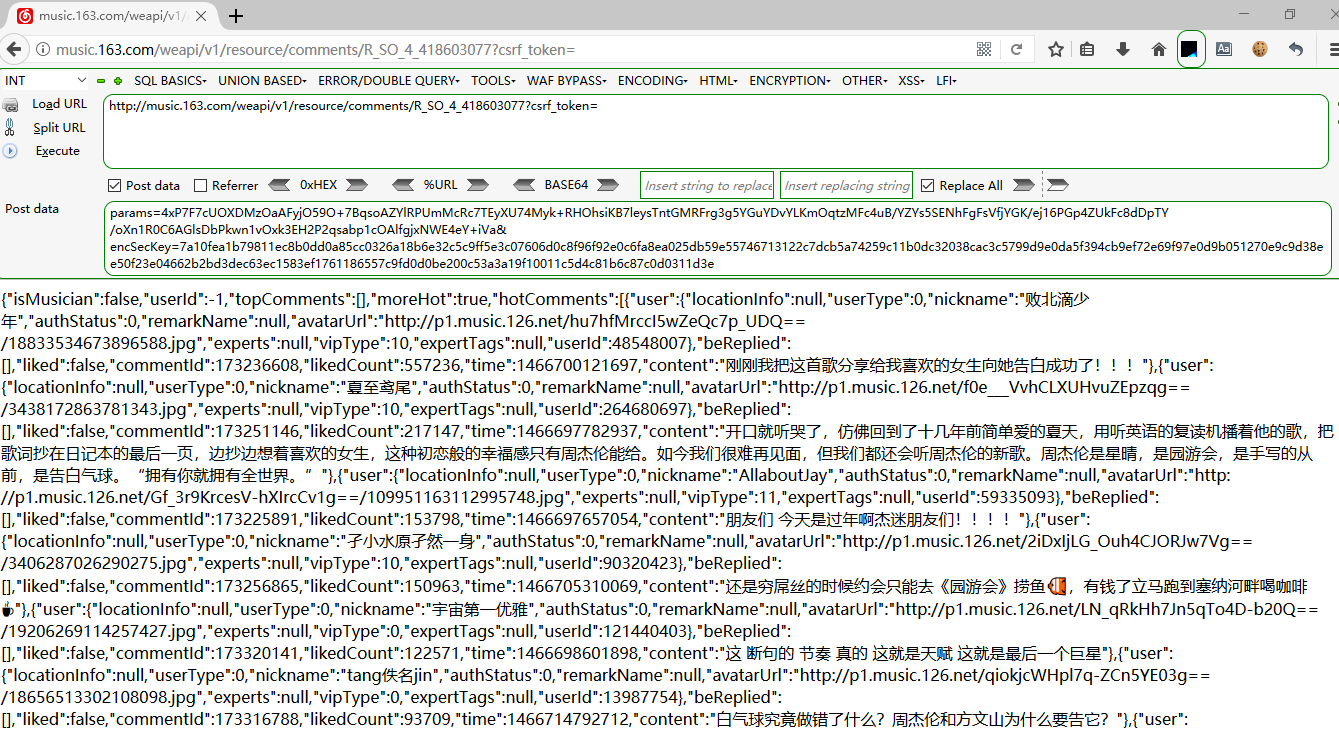
params:

4xP7F7cUOXDMzOaAFyjO59O+7BqsoAZYlRPUmMcRc7TEyXU74Myk+RHOhsiKB7leysTntGMRFrg3g5YGuYDvYLKmOqtzMFc4uB/YZYs5SENhFgFsVfjYGK/ej16PGp4ZUkFc8dDpTY/oXn1R0C6AGlsDbPkwn1vOxk3EH2P2qsabp1cOAlfgjxNWE4eY+iVa

encSecKey:

7a10fea1b79811ec8b0dd0a85cc0326a18b6e32c5c9ff5e3c07606d0c8f96f92e0c6fa8ea025db59e55746713122c7dcb5a74259c11b0dc32038cac3c5799d9e0da5f394cb9ef72e69f97e0d9b051270e9c9d38ee50f23e04662b2bd3dec63ec1583ef1761186557c9fd0d0be200c53a3a19f10011c5d4c81b6c87c0d0311d3e

复制到火狐浏览器查看一下（使用hackbar插件）



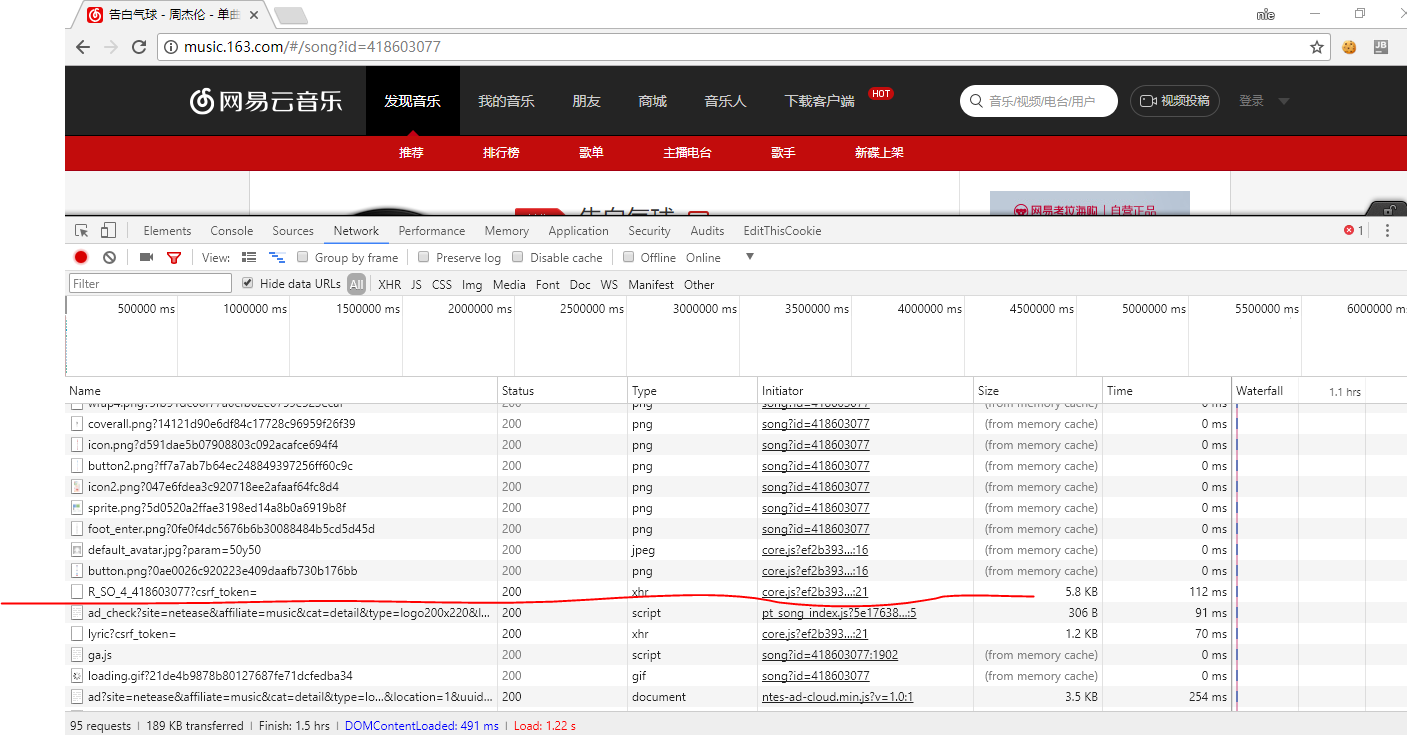
确定这个url是对的

然后就可以尝试用脚本跑出来

但是这里的参数只能跑前面第一页的，不能全部将评论爬出来

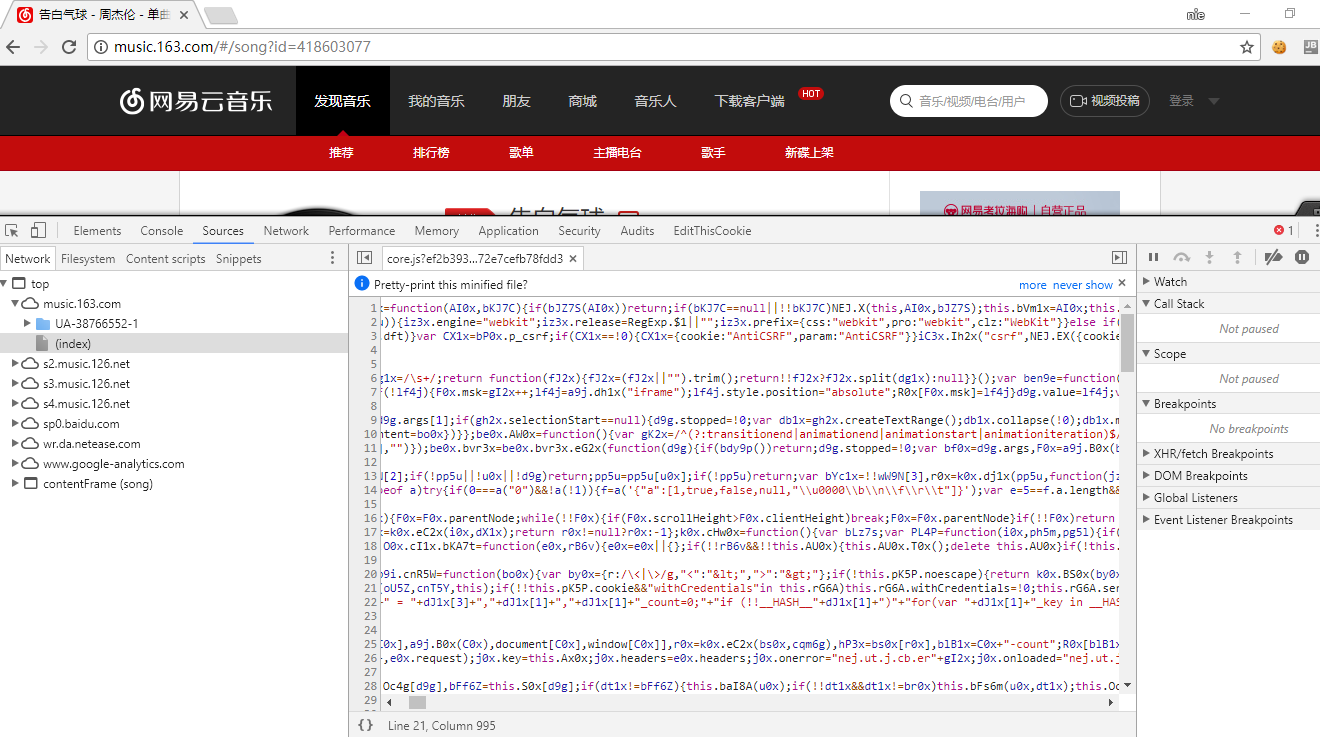
接下来分析参数

继续用Google浏览器



看到使用了js，那就是要分析js文件了

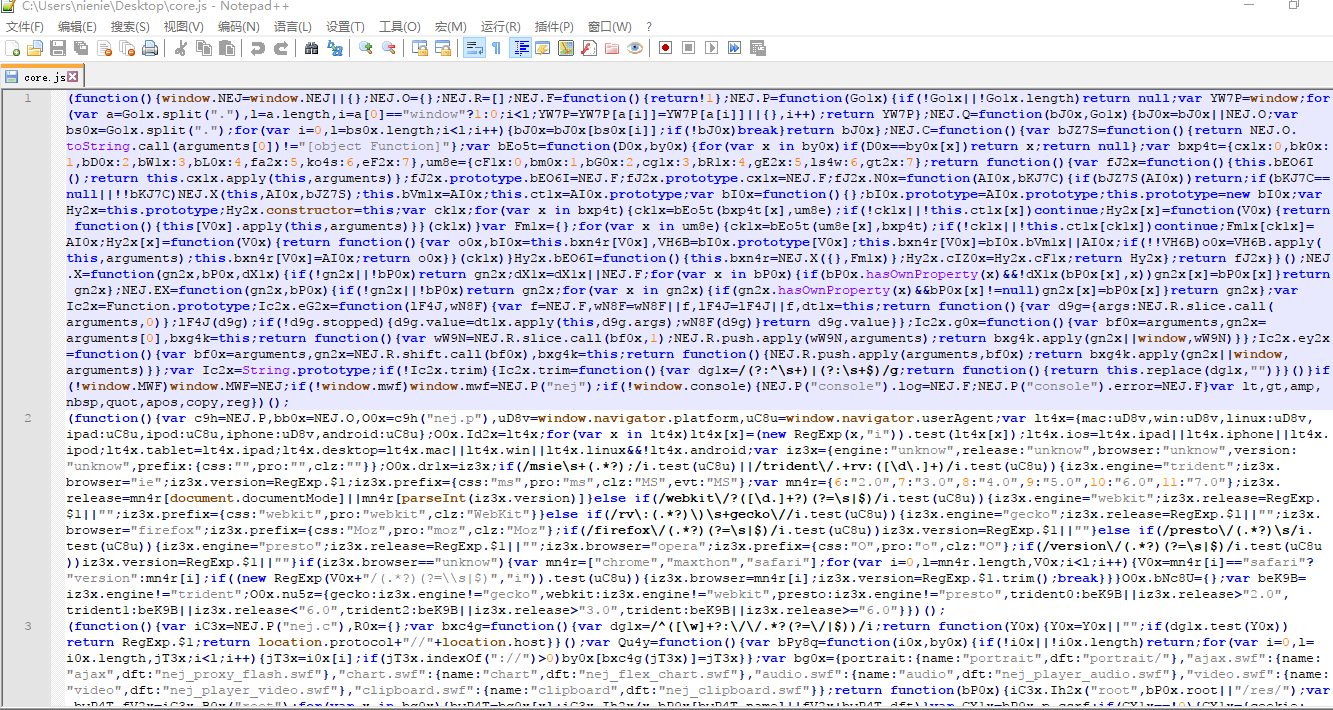
点击这个js文件



右键另存为core.js



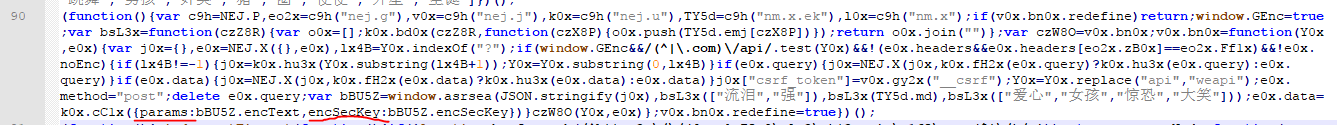
然后用notepad打开



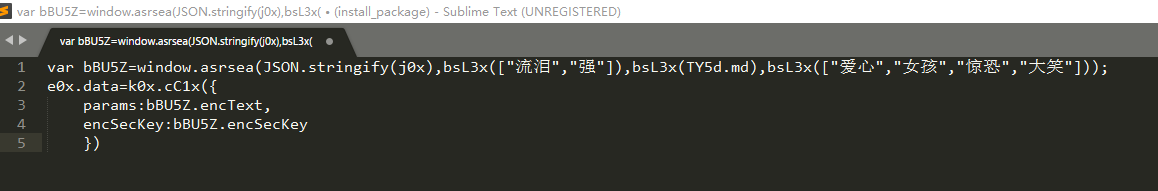
看到这么乱，我们只要查找一下就好了

查找params

不断地找，然后再第90行找到我们所要地params值



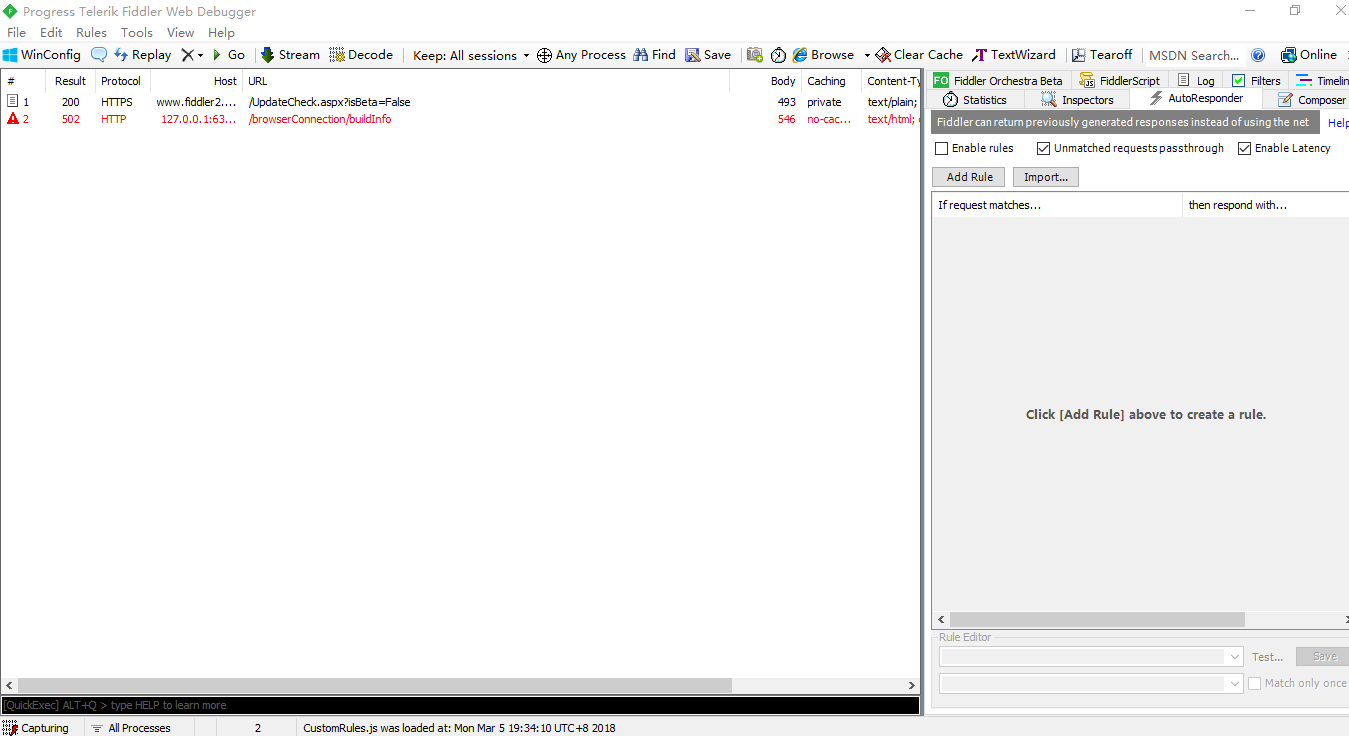
确定这个就是我们要找地js代码，复制出来分析



这里的params和encSecKey都是经过asrsea加密的，然后asrsea里面的内容是什么，和怎样子加密是我们要搞明白的问题

这里有个方便的方法，就是在线调试

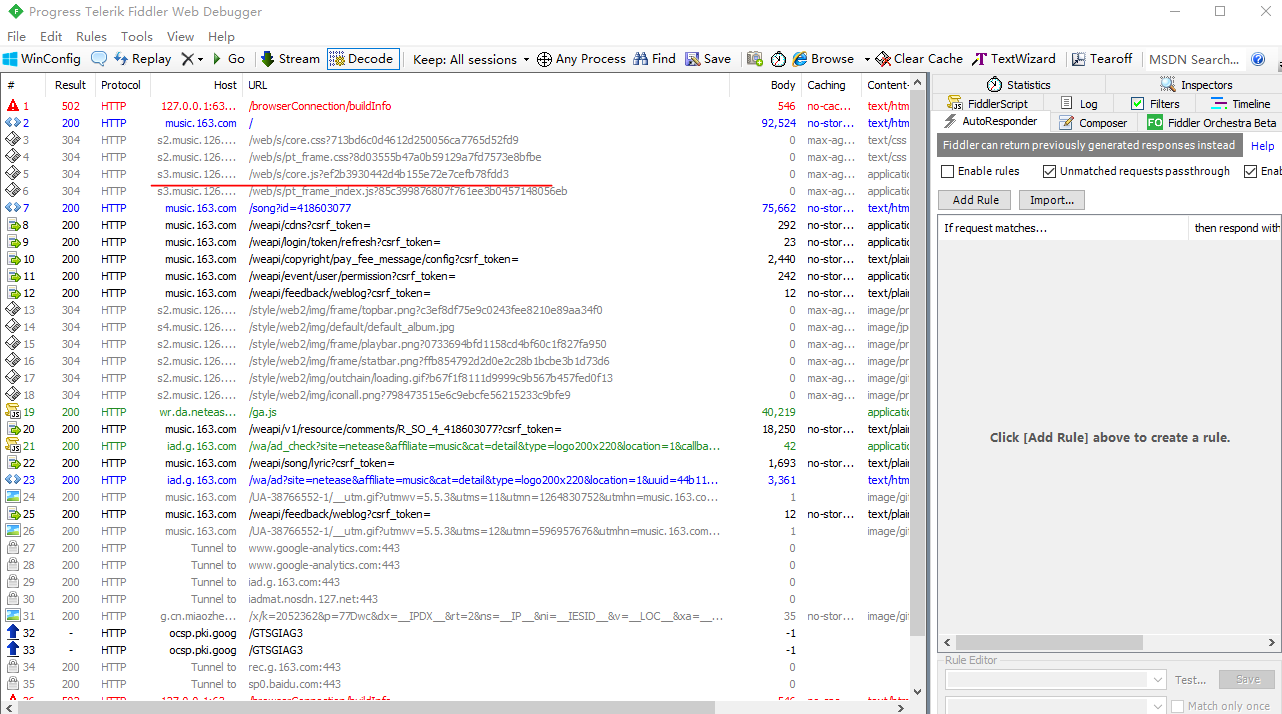
使用Fiddler4



使用火狐浏览器打开那首歌的链接http://music.163.com/#/song?id=418603077

然后设置火狐浏览器的代理为系统代理

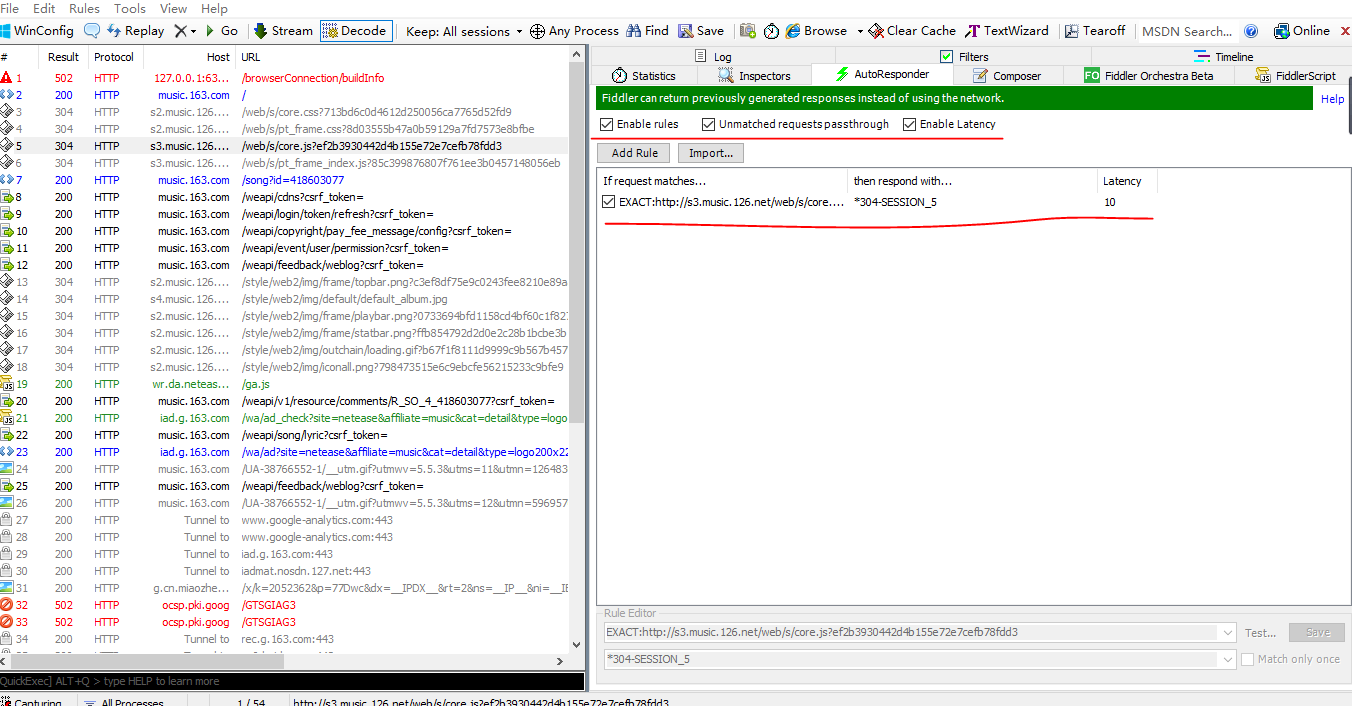
然后刷新那个链接



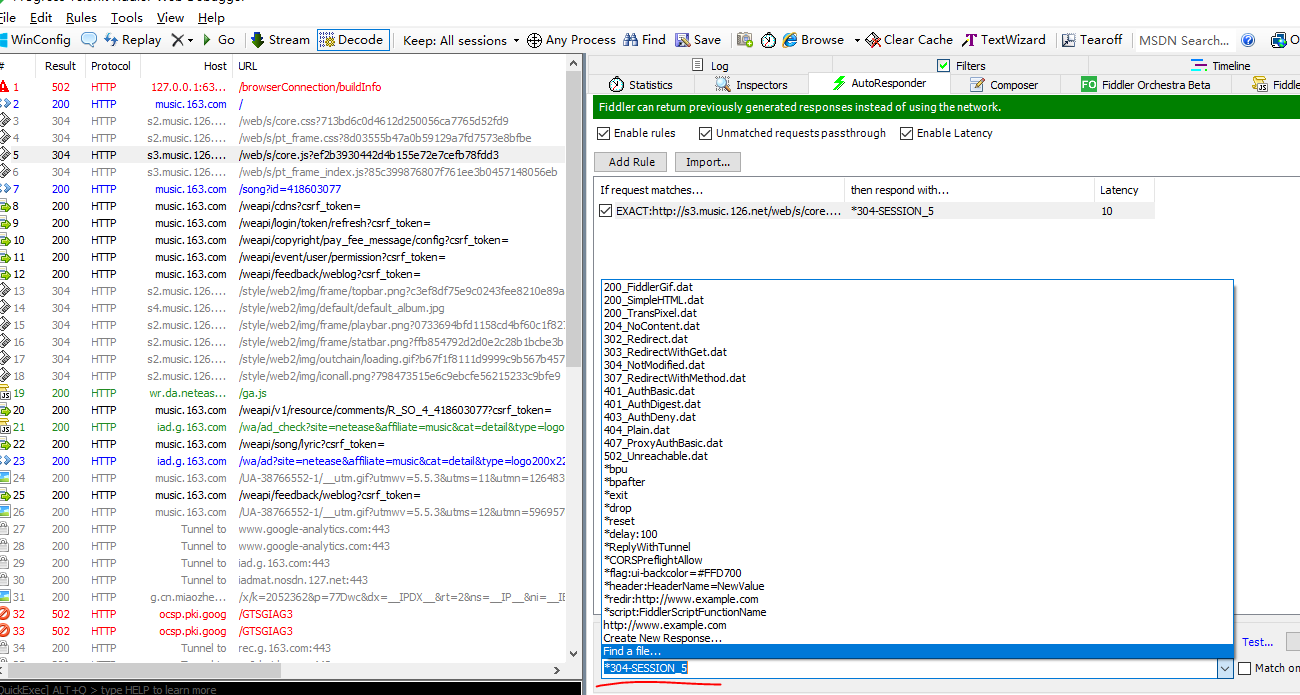
找到这个core.js

点击右边的AutoResponder

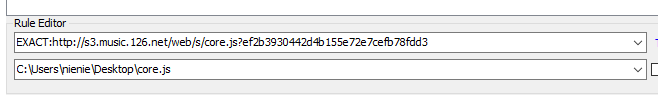
将那个js拉过来



然后选中它，点击下面的rule，替换js为本地的js，然后点击save



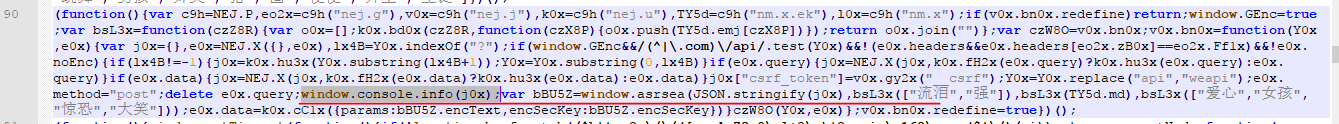
就是这个样子



返回火狐浏览器，点击F12，然后选着控制台，选中日志



然后就用notepad修改js文件测试



在我们需要的代码前面输入window.console.info(j0x);然后保存

刷新火狐浏览器



发现出来了很多数据，这个就是j0x那个参数了

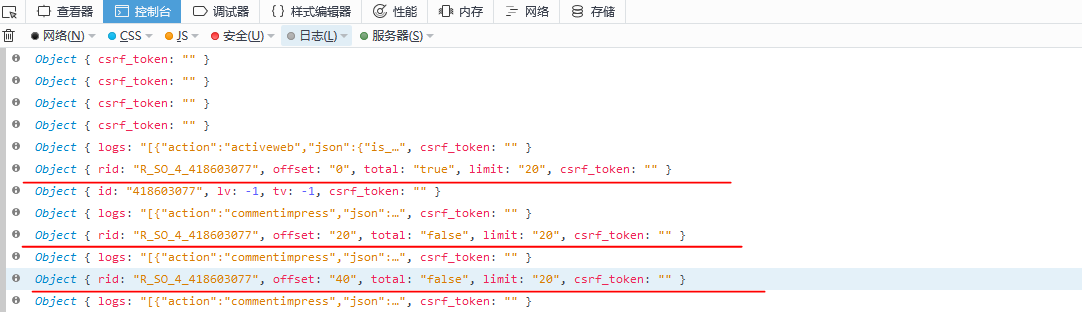
然而{ rid: "R\_SO\_4\_418603077", offset: "0", total: "true", limit: "20", csrf\_token: "" }这个就是我们要的参数，记录下来

尝试点击第二页



又出现了一个json但是里面的参数不一样，offset变成了20，total变成了false

再尝试第三页

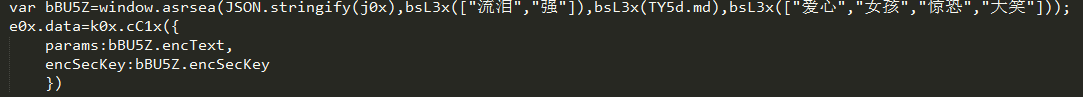


得到结论就是offset就是页数，而且是一个等差数列，offset =（页数-1）\*20

total除了第一页是true之外，其他都是false

Limit猜测是评论的数量，后面验证

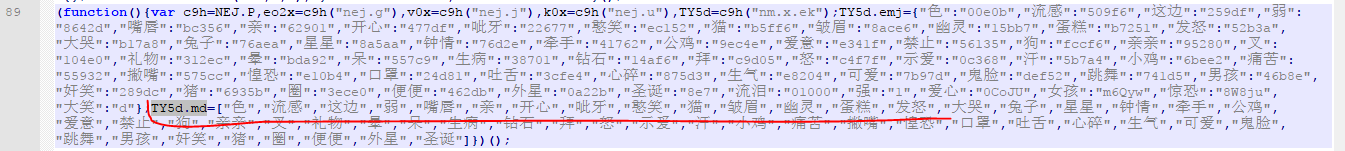
观察



发现那个asrsea的内容又4个，第一个就是刚才的json，第二个是一个固定的值，第三个不知道是不是固定值，第四个也是固定值，现在只要找到第三个，就可以解决了

查看core.js

第89行发现



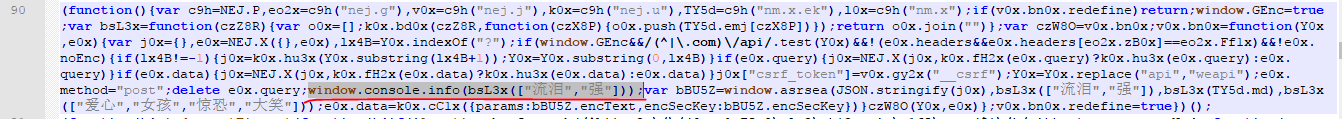
那就是说第三个参数也是个固定值，

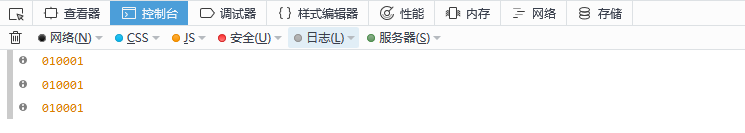
那只要将全部参数都找出来，就可以了

唯一是变量的也就是第一个参数了

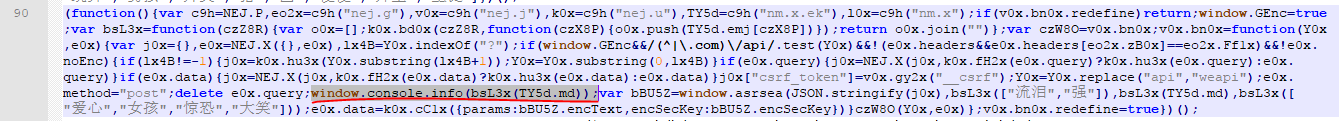
同样方式将那3个参数列出来

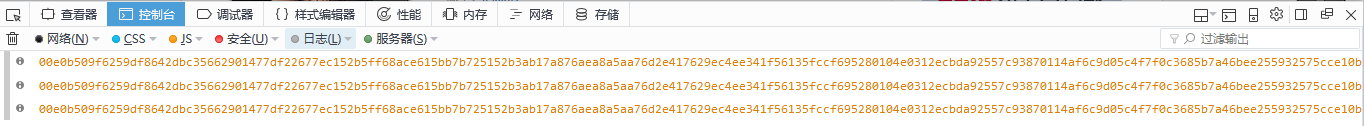
第二个



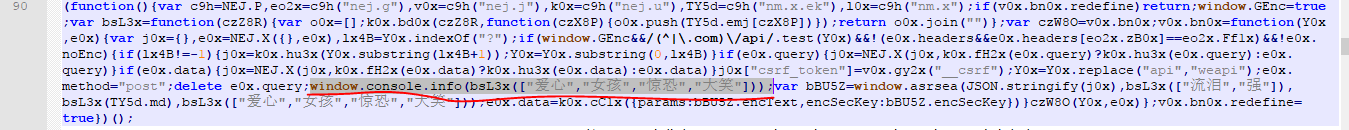


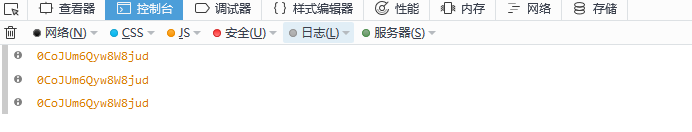
第三个



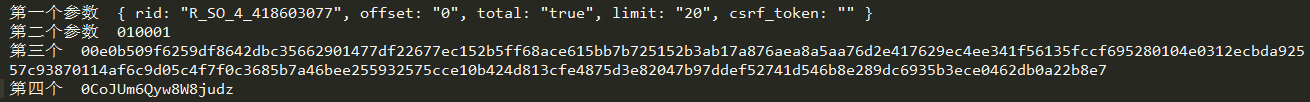


第四个



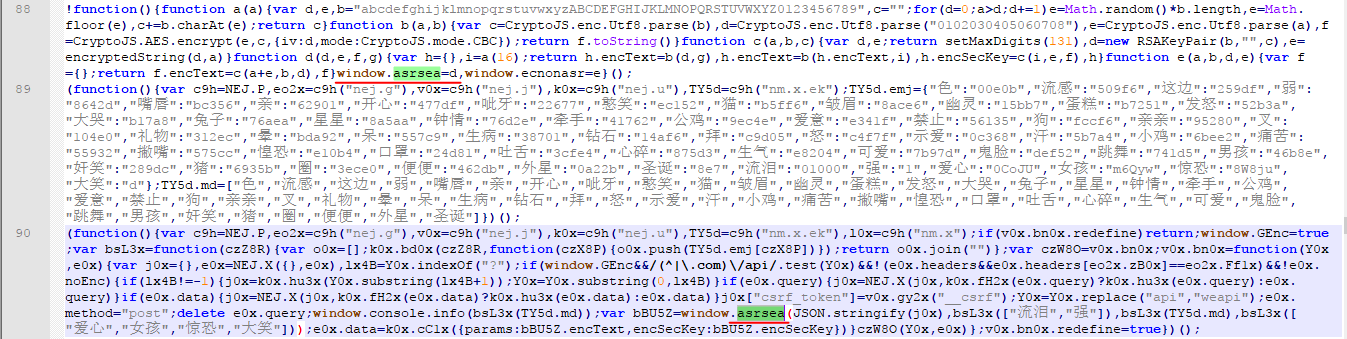


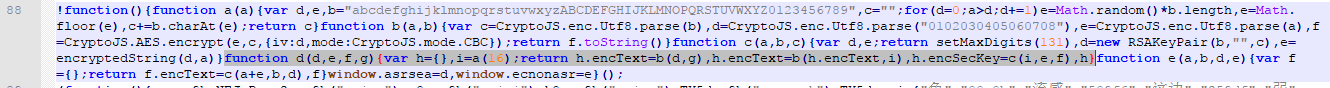
然后就可以得到全部的参数了



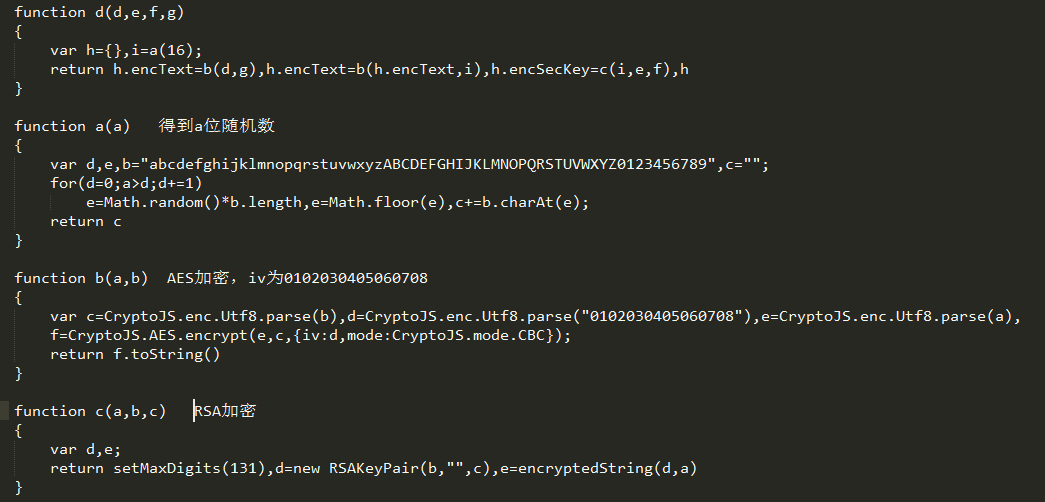
知道了参数然后就是看asrsea的程序了

点击asrsea发现就是等于d方法



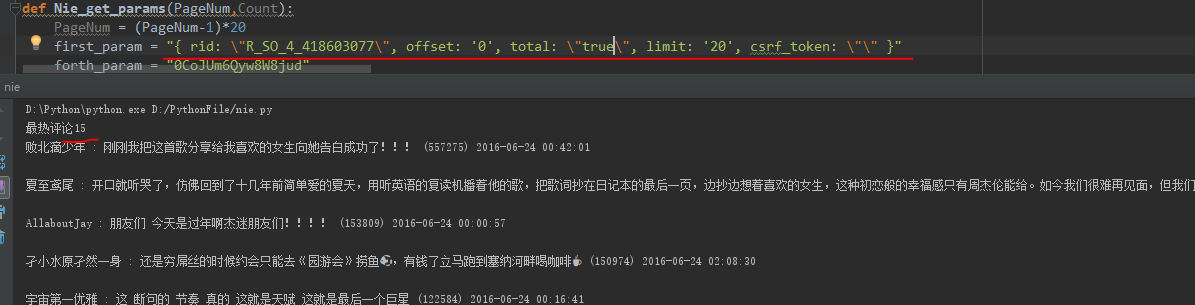


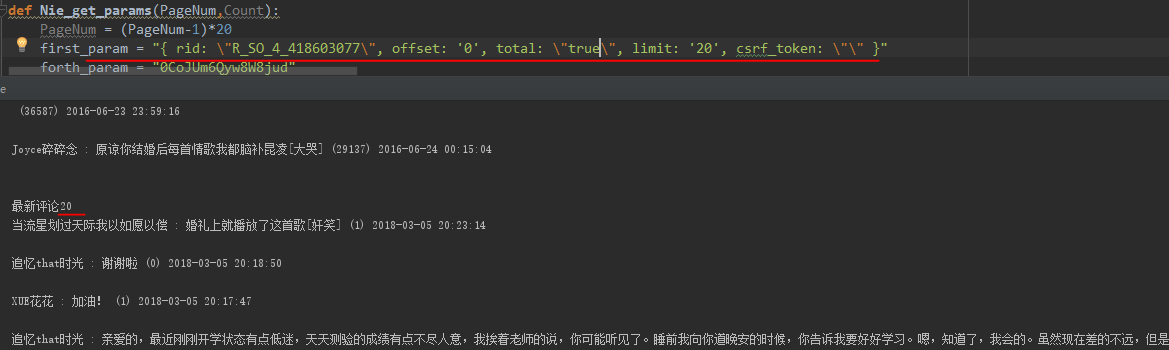
然后找出所有相关的代码程序



然后按照加密，用python写一遍，就可以了

部分代码（完整代码在最后面的链接）

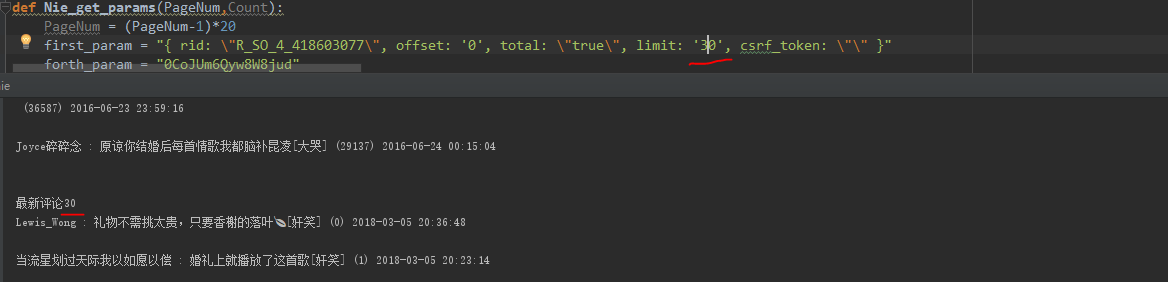




可以看到我们用那些参加加密，然后可以得到评论，最热的有15条，最新的有20条，

尝试修改里面的limit参数

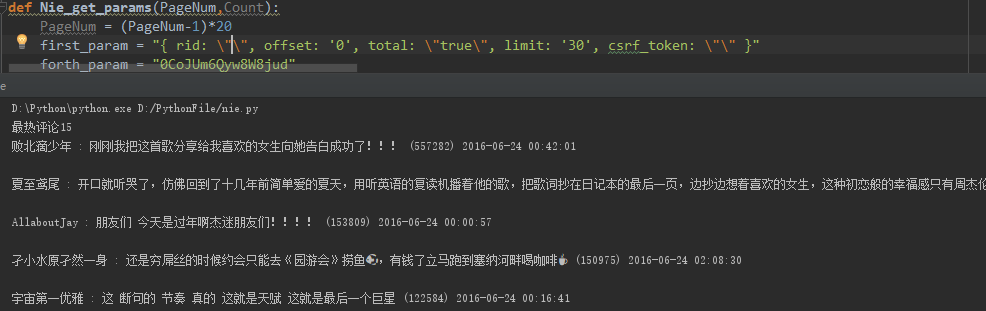
改为30



得到30个最新评论

也就是我们之前的猜测是对的

然后去掉rid尝试

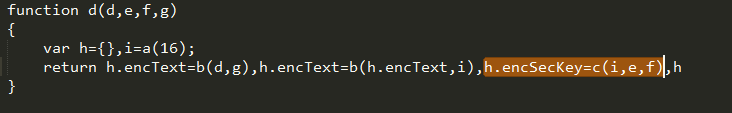


发现还是可以得到那些评论，也就是和那个rid是否有参数没有关系

经过测试，那个limit的最大值为100

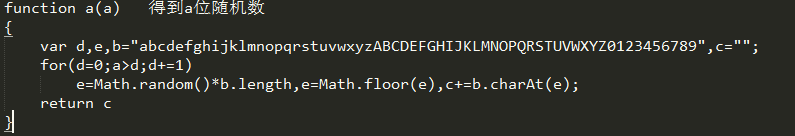
前面一直没有说另外一个参数（encSecKey），因为这个相当于一个常量

观察



发现e为第二个参数，f为第三个参数，也就是这两个都是常量，

i为16为随机数，



这里可以发现是随意产生随机数

两个参数都用到这个i，但是，一旦我们把i就设计为16位随机字母，将它固定了，然后传给两个参数进行加密 ，那么得到的encSecKey就是一个常量，所以我们直接找出一个i就

可以了，用刚才的在线调试，可以找出来。

代码GitHub链接

<https://github.com/niechaojun/NetEaseCloudCrawer>

分析文件在

<https://github.com/niechaojun/Analysis-of-the-NetEase-cloud-comment-parameters>