

抖音爬虫教程

1.模拟器的安装和配置

1. 进入<https://www.ldmng.com/>，下载雷电模拟器并安装
2. 安装完成打开后，在最上方找到“软件设置”，将屏幕调成手机版



3. 在电脑端启动cmd（命令提示符），输入ipconfig获取本机ip，复制

```
选择命令提示符

无线局域网适配器 本地连接* 2:

    媒体状态 . . . . . : 媒体已断开连接
    连接特定的 DNS 后缀 . . . . . :

以太网适配器 以太网:

    连接特定的 DNS 后缀 . . . . . :
    IPv6 地址 . . . . . : 240e:3a1:6840:55e0:7cc0:975c:4a0c:bb49
    临时 IPv6 地址. . . . . : 240e:3a1:6840:55e0:cd4c:e2b4:abaf:b8db
    本地链接 IPv6 地址. . . . . : fe80::7cc0:975c:4a0c:bb49%14
    IPv4 地址 . . . . . : 192.168.1.2
    子网掩码 . . . . . : 255.255.255.0
    默认网关. . . . . : fe80::1%14
                        192.168.1.1

以太网适配器 蓝牙网络连接:

    媒体状态 . . . . . : 媒体已断开连接
    连接特定的 DNS 后缀 . . . . . :

C:\Users\ThinkPad>
```

4. 在模拟器中进入设置，长按第一个无线，点击修改网络



WLAN



开启



what

已连接 (通过 WLAN 助手)



按键



加量



减量



全屏



截图



多开



安装



设置



更多





WLAN



开启



what
已连接

what

取消保存网络

修改网络



按键



加量



减量



全屏



截图



多开



安装



设置



更多



5. 将代理选项切换成手动，粘贴你刚才复制的ip，保存（服务器端口设为8888，等会有用）

what

☒ 高级选项

代理
手动 ▼

浏览器会使用 HTTP 代理，但其他应用可能不会使用。

代理服务器主机名

192.168.1.2

代理服务器端口

8888

对以下网址不使用代理

example.com,mycomp.test.com,l

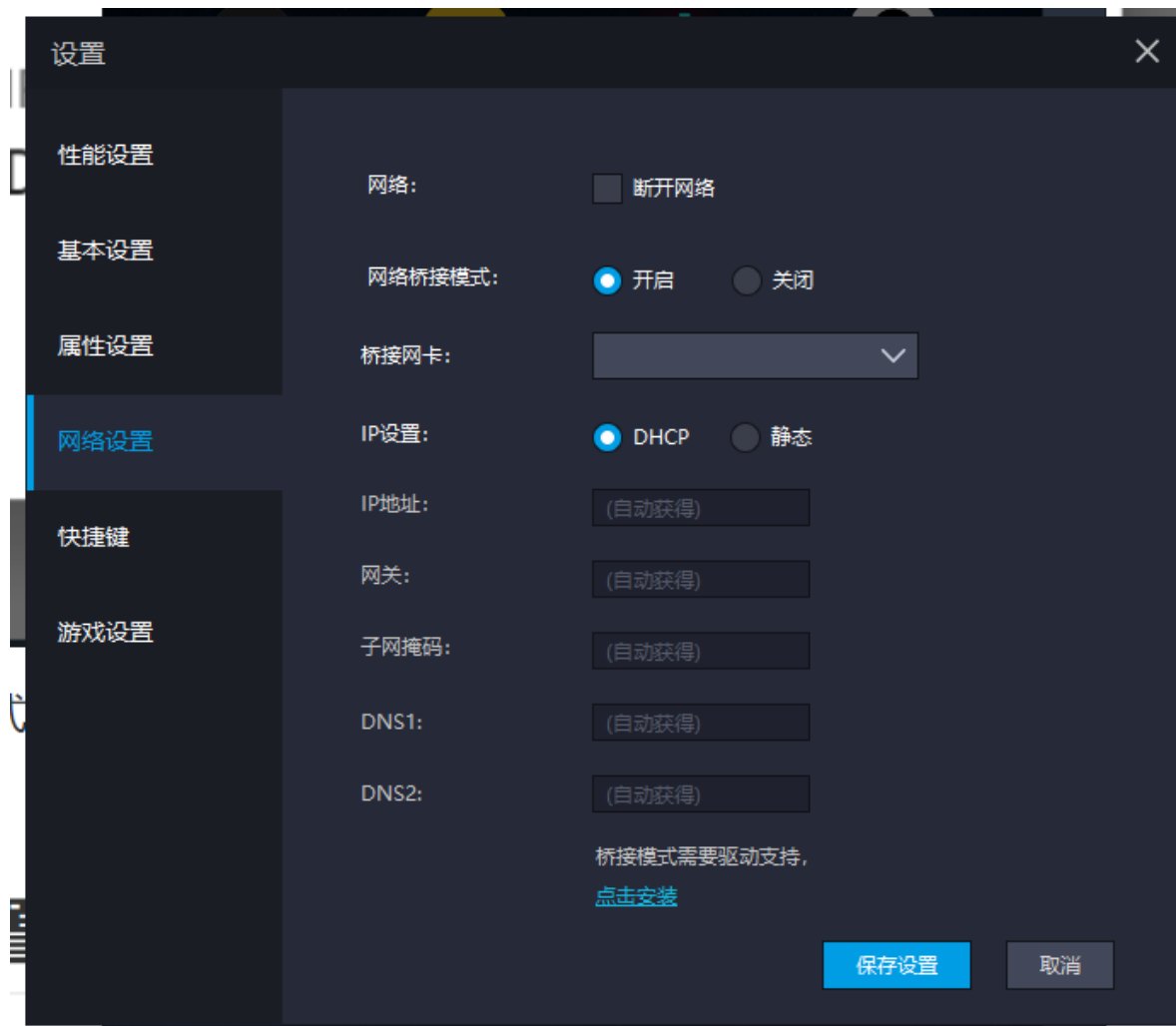
IP 设置

DHCP ▼

取消

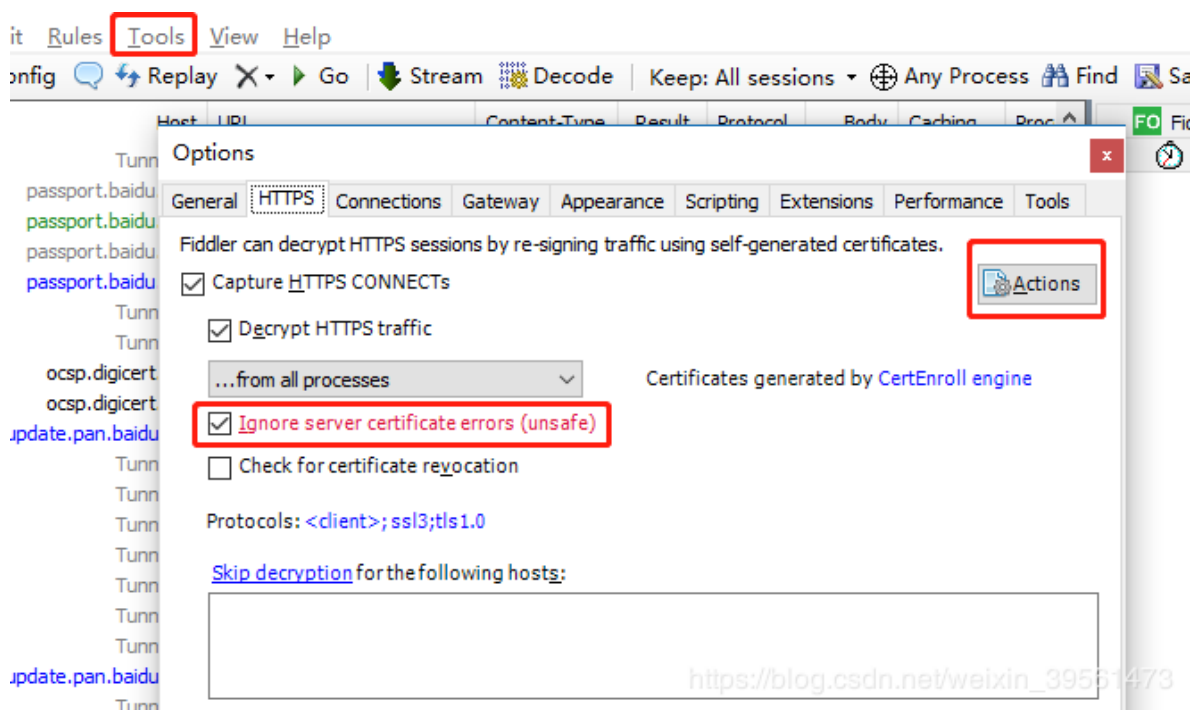
保存

6. 在设置中开启桥接模式，需要在下面点击安装驱动支持。



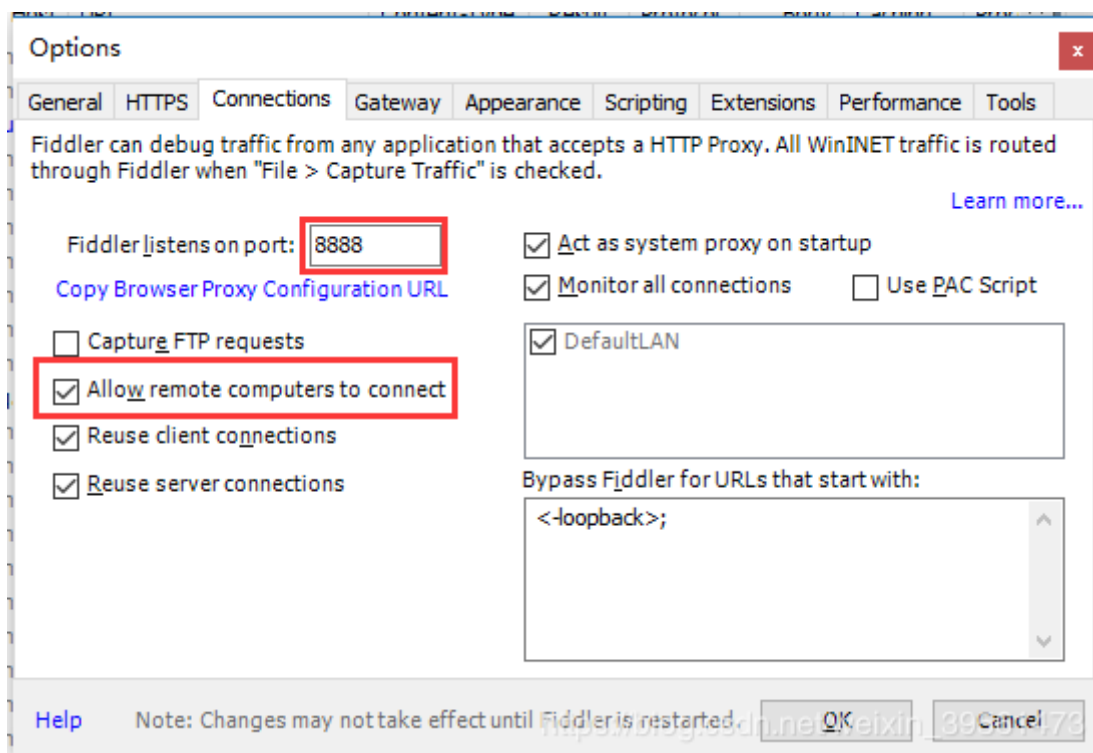
2. fiddler的配置使用

1. 安装<https://www.telerik.com/download/fiddler>
2. 在tools中的options中，按照图中勾选后点击Actions



3. 配置远程链接：

选择允许监控远程链接，端口可以随意设置，只要别重复就行，默认8888（照应上面在模拟器设置的8888）



4. 重启fiddler，一定要重启！

3. 模拟器链接fiddler

代理设置好后，在浏览器中输入你设置的ip：端口，例如192.168.3.10(这里是您刚才的 ip地址)：8888，就会打开fiddler的页面。然后点击**fiddlerRoot certificate**安装证书，要不手机会认为环境不安全。



http://192.168.3.10:8888/



Fiddler Echo Service

GET / HTTP/1.1
Host: 192.168.3.10:8888
Proxy-Connection: keep-alive
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Linux; Android 5.1.1; V1916A Build/UM49T; wv) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0 Chrome/52.0.2743.100 Mobile Safari/537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Encoding: gzip, deflate
Accept-Language: zh-CN,zh-US;q=0.8
X-Requested-With: com.android.browser

This page returned a **HTTP/200** response

Originating Process Information:

Idboxheadless:20120

- To configure Fiddler as a reverse proxy instead of seeing this page, see [Reverse Proxy Setup](#)
- You can download the [FiddlerRoot certificate](#)



按键



加量



减量



全屏



截图



多开



安装



设置



更多



下载好后点任务栏，安装一下，按照模拟器步骤做 就好了

4.抖音下载

1. 经过上面这些步骤，你的模拟器就能上网了
2. 首先在浏览器输入 baidu.com进入百度（默认谷歌，因为一些众所周知的原因上不了）（忽略图片最下面一行.....）



https://m.baidu.com/?fro



抢寄快递神券

¥68



网页版



百度一下



百度APP

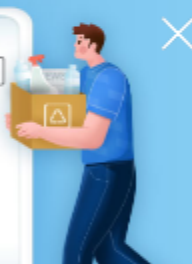
ziroom自如

租房生活

垃圾分类 自如应对

来百度APP体验垃圾分类神器

立即体验



习主席讲话让世界感受中国情谊与担当

置顶 新华社客户端

中非团结抗疫，这些承诺习近平再强调

置顶 人民网



新发地619亩临时交易区
已投入使用

环球网

世界上交配时间最长的动物，时间竟长达

3. 下载抖音安装

4. 打开后，就和你手机上看着没啥区别了



按键



加量



减量



全屏



截图



多开



安装



设置



更多



游戏名：欢乐贪吃龙

点击左下角进入游戏

宝石兑换码

100100	100110
100111	111111
111112	222222
333333	A798798

龙皮肤兑换码

520520
666888
168168
999999
618618

龙兑换码

欢乐贪吃龙 #nzhuermiao

@猫小妖游戏 @aojiayou

相信很多小伙伴都已经拿到了

#欢乐贪吃龙 别再被骗了，无限钻石是假的，但是这个隐藏彩蛋是真的 #学生党 #00后 #抖音小游戏

🎵 - 苏阿三 @苏阿三创作的

📝 喜欢就要说出来

@



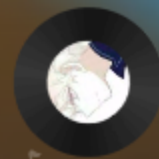
2.7w



1611

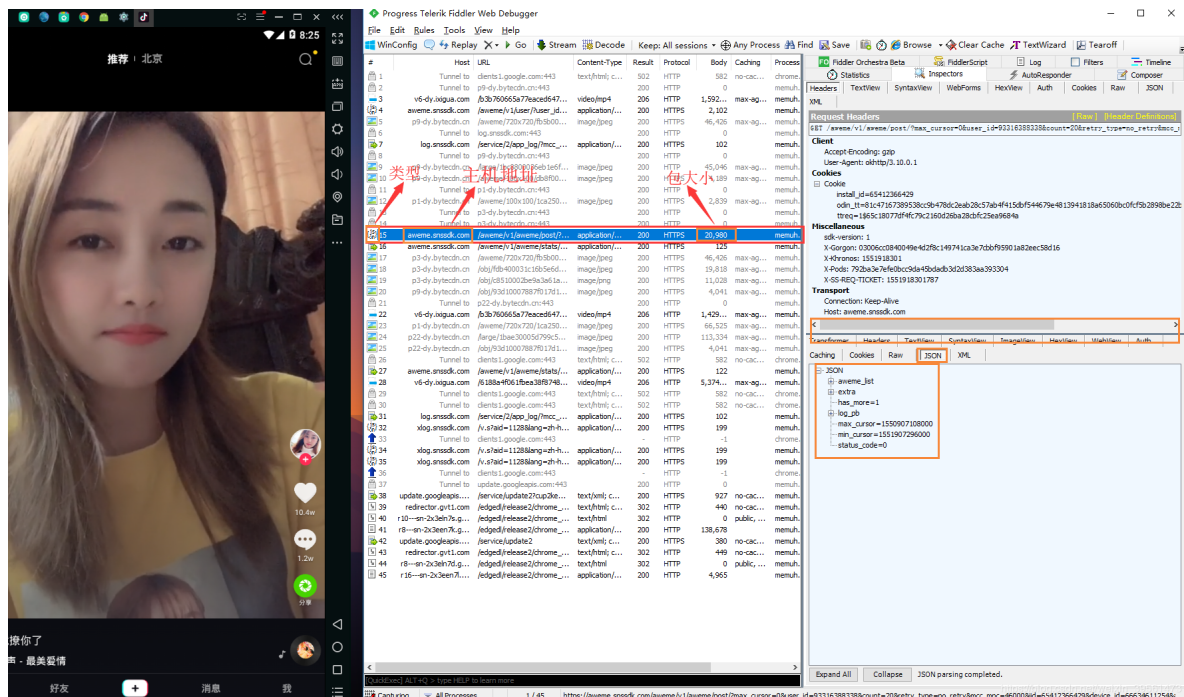


分享

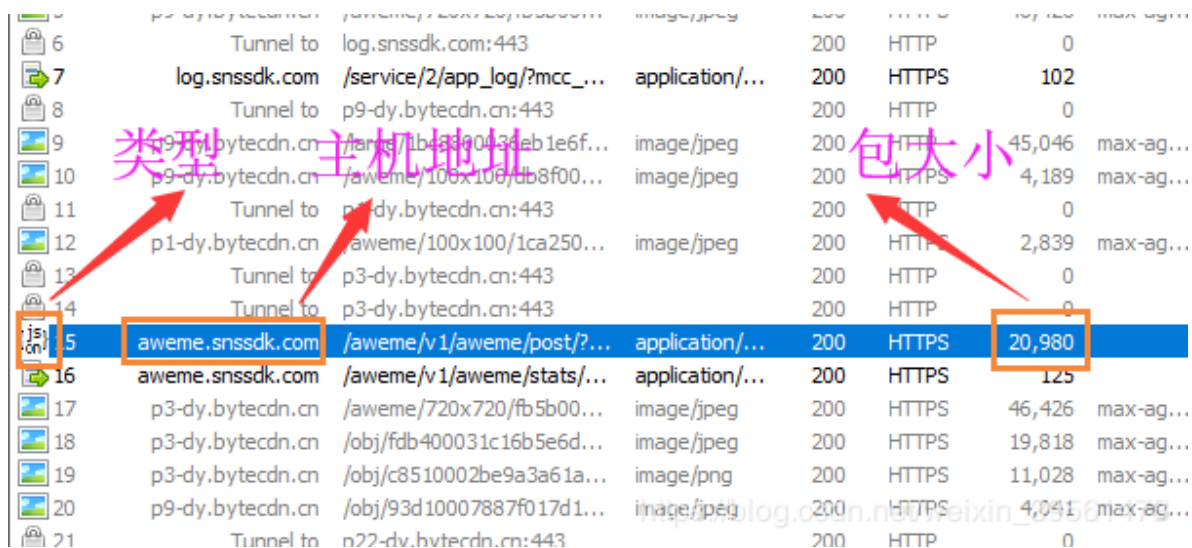


5.fiddler抓包

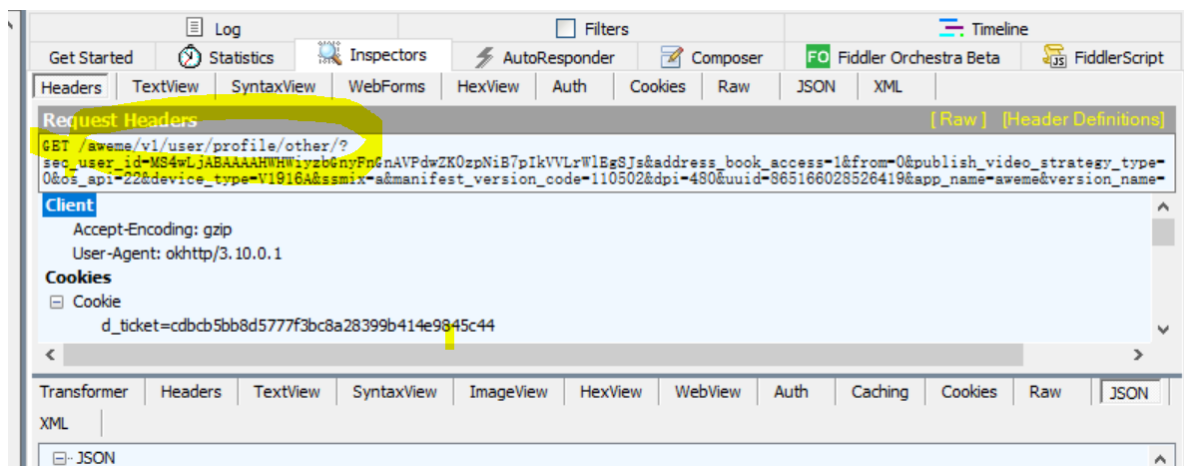
1. 在浏览了一些抖音视频后，你的fiddler就会出现很多包
2. 观察出现的所有包。



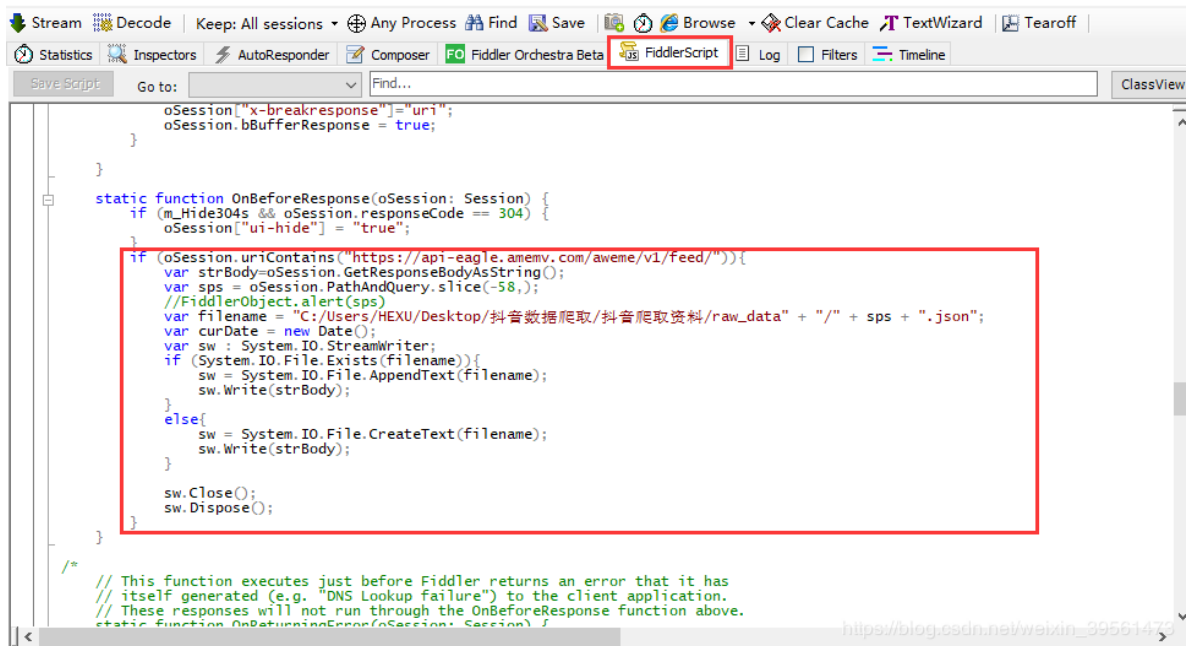
3. 其中有个包，包类型为json (json就是网页返回的数据，具体百度)，主机地址如图。这个包里一般包含着相关的信息。



4. 每个包都有一个Request Headers，其中形如下图的是个人资料包。



5. 我们使用fiddler自带的脚本，在里面添加规则，当指定的json包刷出来后自动保存json包。



6. 我们要抓取个人资料包，代码如下。

```
if
(oSession.uriContains("https://aweme.snssdk.com/aweme/v1/user/profile/other/")){
    var strBody=oSession.GetResponseBodyAsString();
    var sps = oSession.PathAndQuery.slice(-19,);
    //FiddlerObject.alert(sps)
    var date = new Date();
    var y = date.getFullYear();
    var m = date.getMonth() + 1;
    m = m < 10 ? ('0' + m) : m;
    var d = date.getDate();
    d = d < 10 ? ('0' + d) : d;
    var h = date.getHours();
    var minute = date.getMinutes();
    minute = minute < 10 ? ('0' + minute) : minute;
    var second= date.getSeconds();
    second = minute < 10 ? ('0' + second) : second;
    var str = m+"-"+d+"-"+h+minute+second;

    var filename = "C:/Users/ThinkPad/Desktop/data" + "/" + str +
    ".json";

    var curDate = new Date();
    var sw : System.IO.StreamWriter;
    if (System.IO.File.Exists(filename)){
        sw = System.IO.File.AppendText(filename);
        sw.Write(strBody);
    }
    else{
        sw = System.IO.File.CreateText(filename);
        sw.Write(strBody);
    }

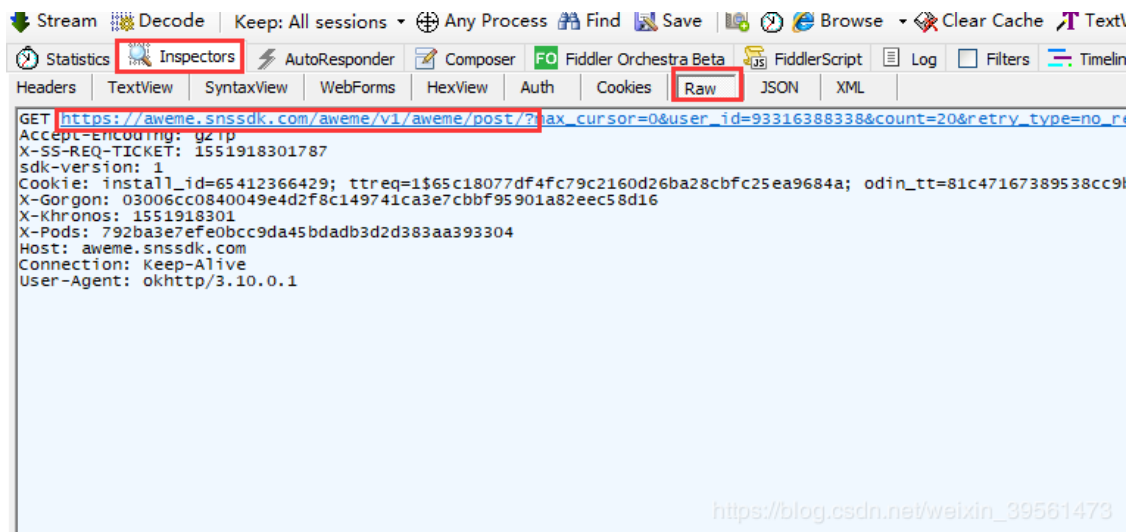
    sw.Close();
    sw.Dispose();
}
```


7. 这个脚本有两点需要修改的:

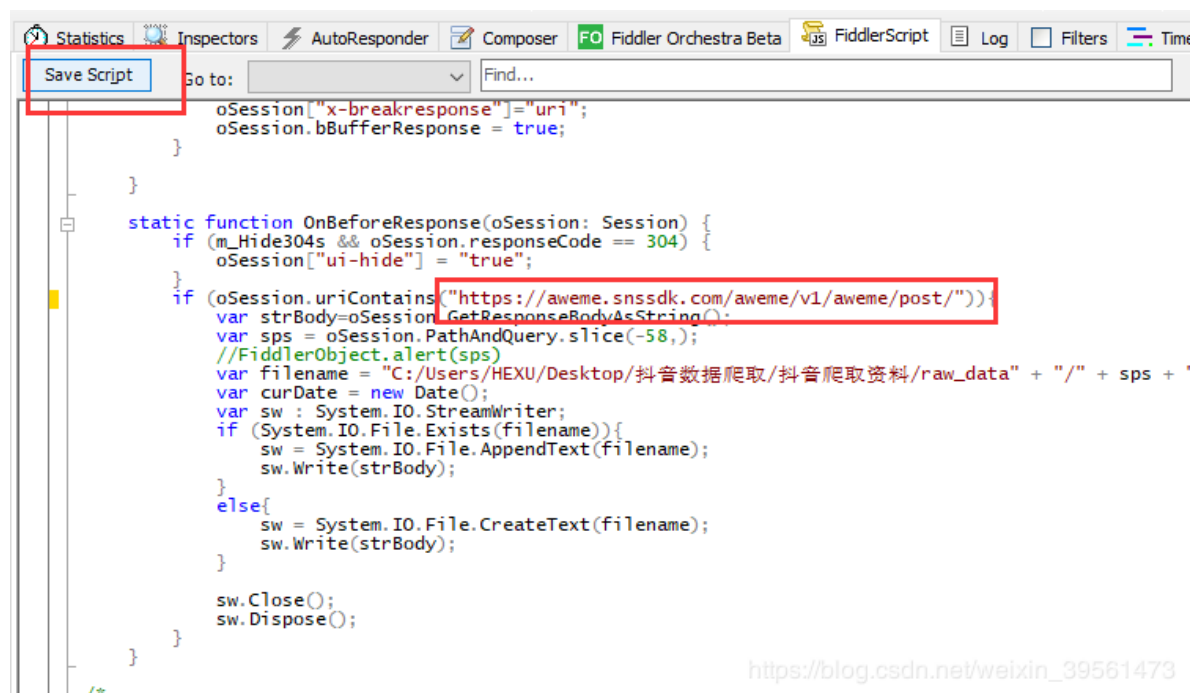
(1) 第一行的网址:

这个是从包的url中摘出来的, 抖音会时不时更新这个url, 所以不能用了也要去更新:

比如现在的已经和昨天不同了, 记着修改。



(2) 路径, 那个是我设置json包保存的地址, 自己一定要去修改, 并创建文件夹, 修改完记着点保存。



8. 修改完毕, 一定要点“save script”保存

9. 打开设置好模拟器和脚本后, 等待一会, 就可以看到文件夹中保存的包了:

