python 爬取手机 app

App 的爬取相比 web 端爬取更加容易，反爬虫能力没有那么强，而且数据大多以 json 形式传输，解析更加简单。

在 web 端，可以通过浏览器的开发这工具监听到各个网络请求和响应过程，**在 App 端如果想要查看这些内容就需要借助抓包软件。**

原理：通过设置代理的方式确保手机和 pc 处于同一个局域网内，将手机处于抓包软件的监听之下，这样 app 发给服务器的数据包和服务器返回的数据包都要经由代理服务器转发，抓包软件便可以看到 App 运行过程中的请求和响应了，如果这些请求的 url，参数是有规律的，就可以总结出规律直接用程序模拟爬取。

Fiddler 是最强大最好用的抓包工具之一， 使用 Fiddler 对开发会有很大的帮助。

Fiddler 的基本介绍

官方网站: https://www.telerik.com/fiddler

Fiddler 是最强大最好用的 Web 调试工具之一

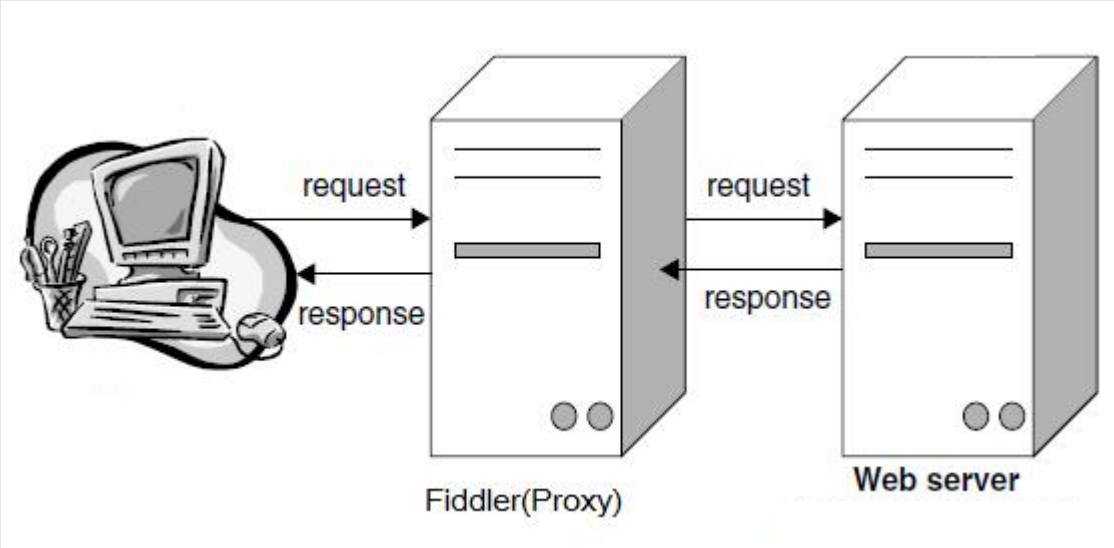
- 它能记录所有客户端和服务器的 http 和 https 请求
- 允许监视，设置断点，甚至修改输入输出数据
- Fiddler 包含了一个强大的基于事件脚本的子系统，并且能使用.net 语言进行扩展

你对 HTTP 协议越了解， 你就能越掌握 Fiddler 的使用方法. 你越使用 Fiddler,就越能帮助你了解 HTTP 协议.

Fiddler 的工作原理

Fiddler 是以代理 web 服务器的形式工作的，它使用代理地址:127.0.0.1，端口:8888.

Fiddler 会自动设置代理，退出的时候它会自动注销代理，这样就不会影响别的程序。不过如果 Fiddler 非正常退出，这时候因为 Fiddler 没有自动注销，会造成网页无法访问。解决的办法是重新启动下 Fiddler.

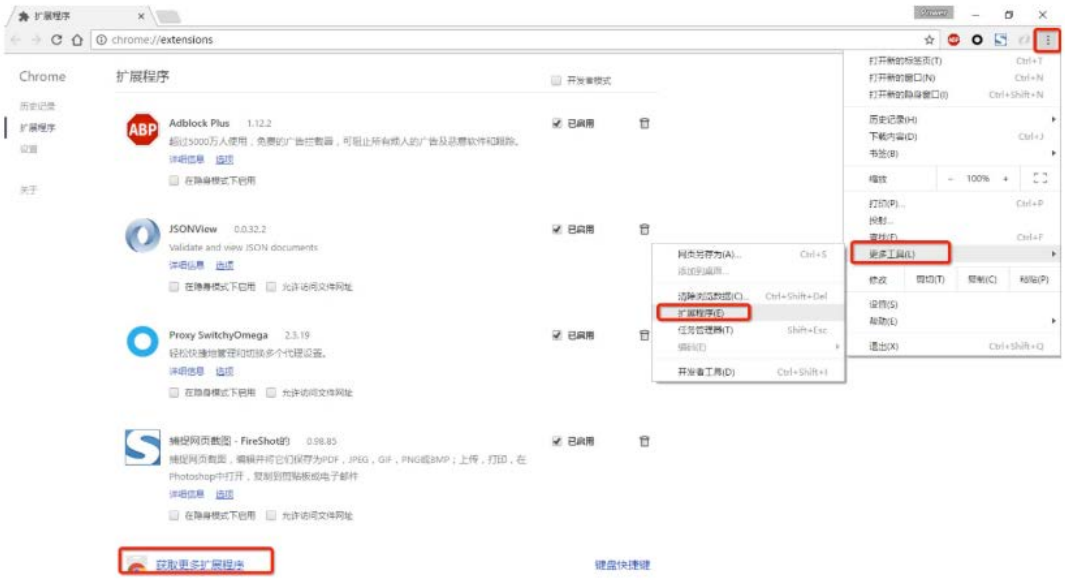


同类的其它工具

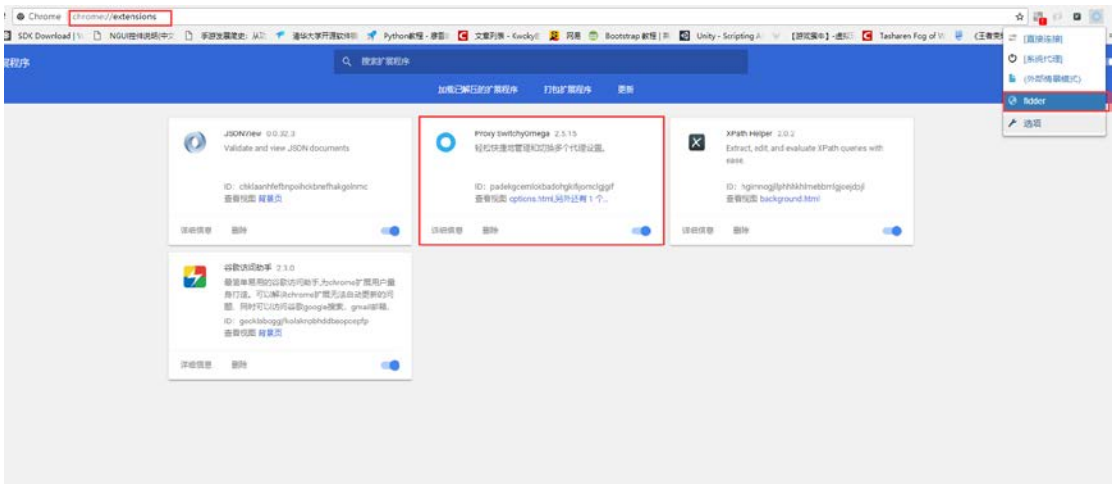
同类的工具有: httpwatch, firebug, wireshark

Fiddler 如何捕获 Chrome 的会话

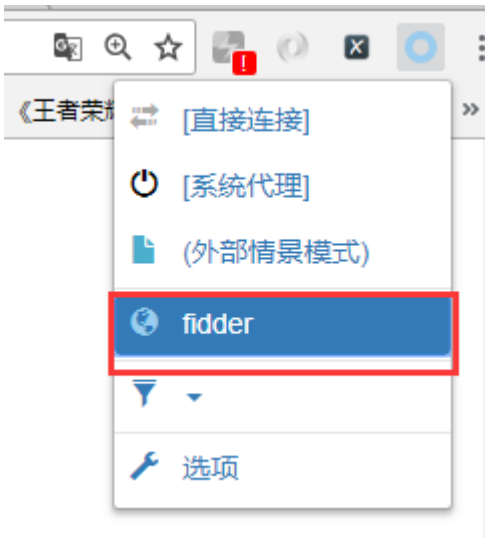
安装 SwitchyOmega 代理管理 Chrome 浏览器插件



如图所示，设置代理服务器为 127.0.0.1:8888

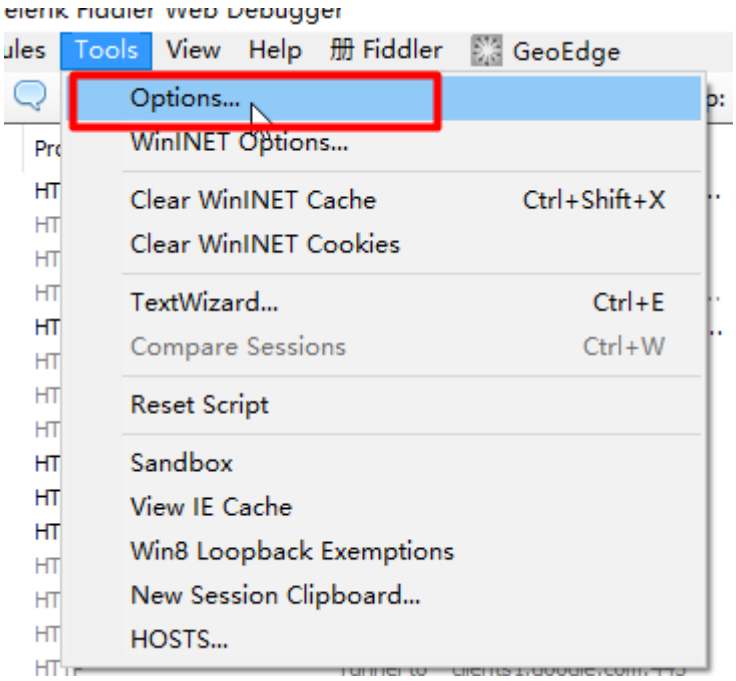


通过浏览器插件切换为设置好的代理。



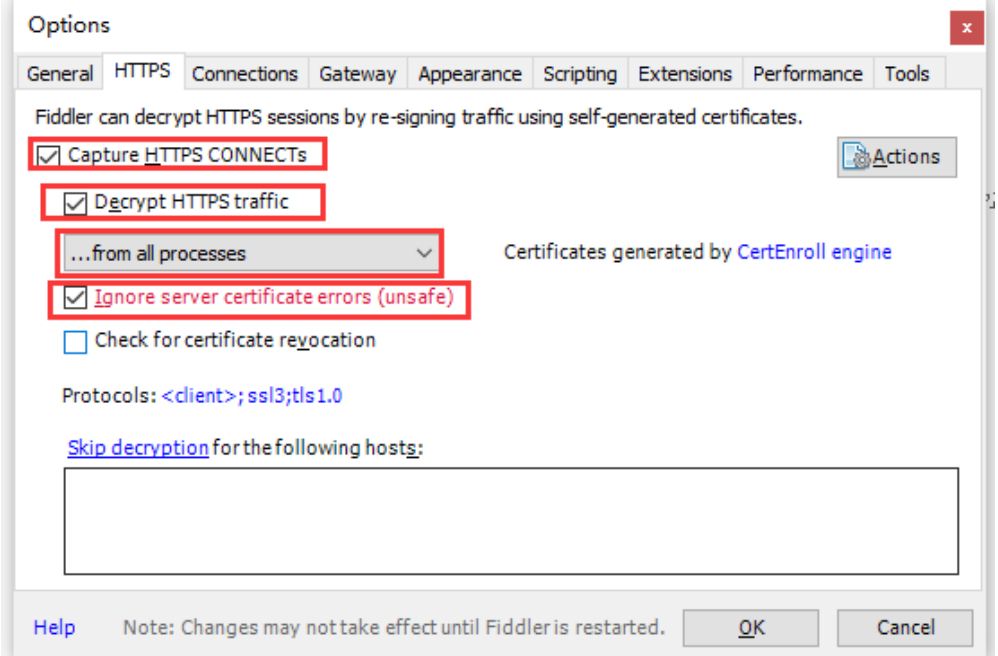
Fiddler 捕获 HTTPS 会话设置

默认下，Fiddler 不会捕获 HTTPS 会话，需要你设置下，打开 Fiddler Tool-> Options->HTTPS tab

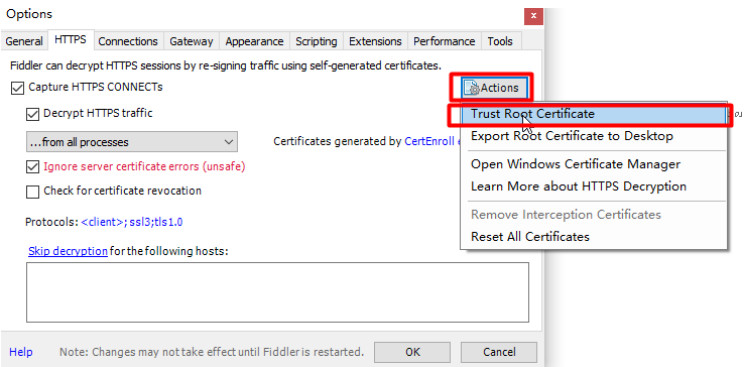


对 Fiddler 进行设置：

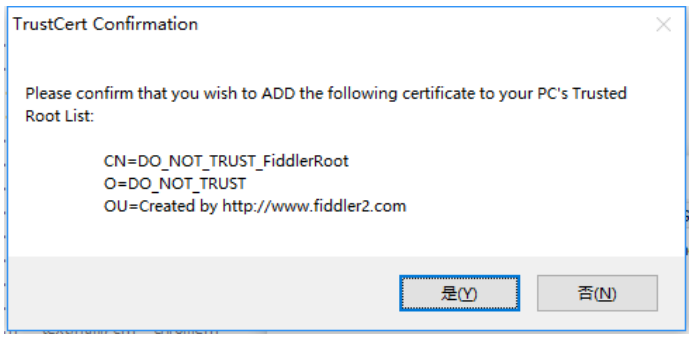
- 打开工具栏->Tools->Fiddler Options->HTTPS，
- 选中 Capture HTTPS CONNECTs (捕捉 HTTPS 连接)，
- 选中 Decrypt HTTPS traffic（解密 HTTPS 通信）
- 另外我们要用 Fiddler 获取本机所有进程的 HTTPS 请求，所以中间的下拉菜单中选中...from all processes （从所有进程）
- 选中下方 Ignore server certificate errors（忽略服务器证书错误）



为 Fiddler 配置 Windows 信任这个根证书解决安全警告：Trust Root Certificate（受信任的根证书）。



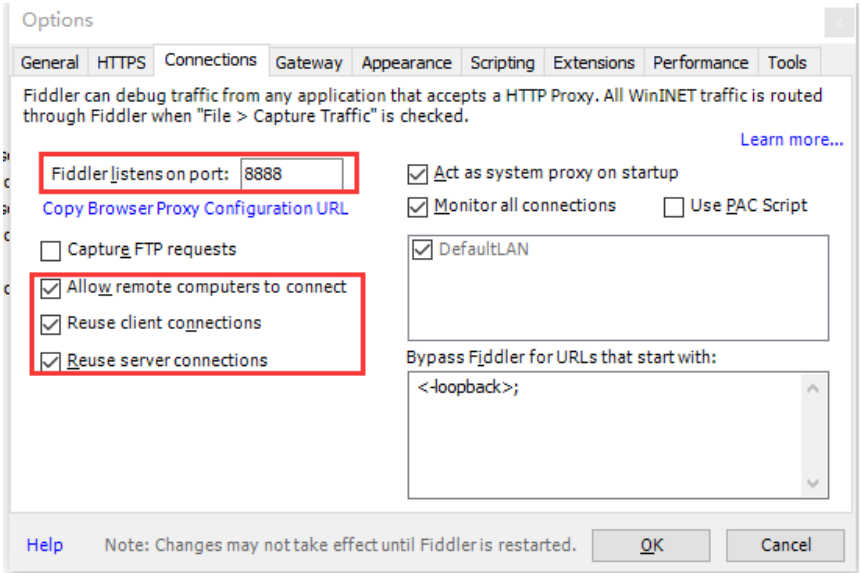
弹出如下的对话框，点击"YES"



点击"Yes" 后，就设置好了

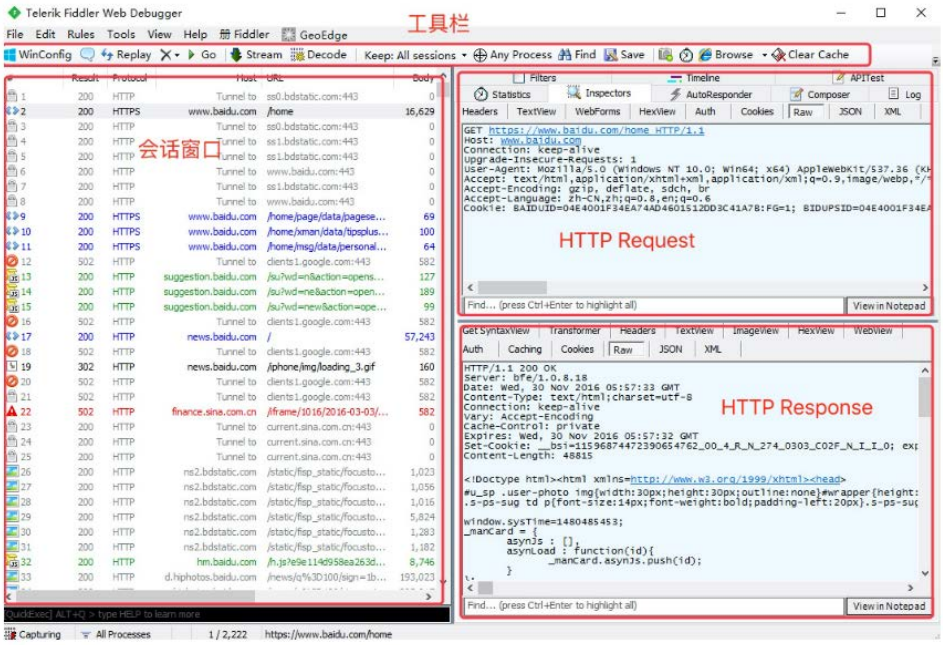
Fiddler connections 设置

点击 Connections，将 Fiddler listens on port 设为 8888，勾选 Allow remote computers to connect

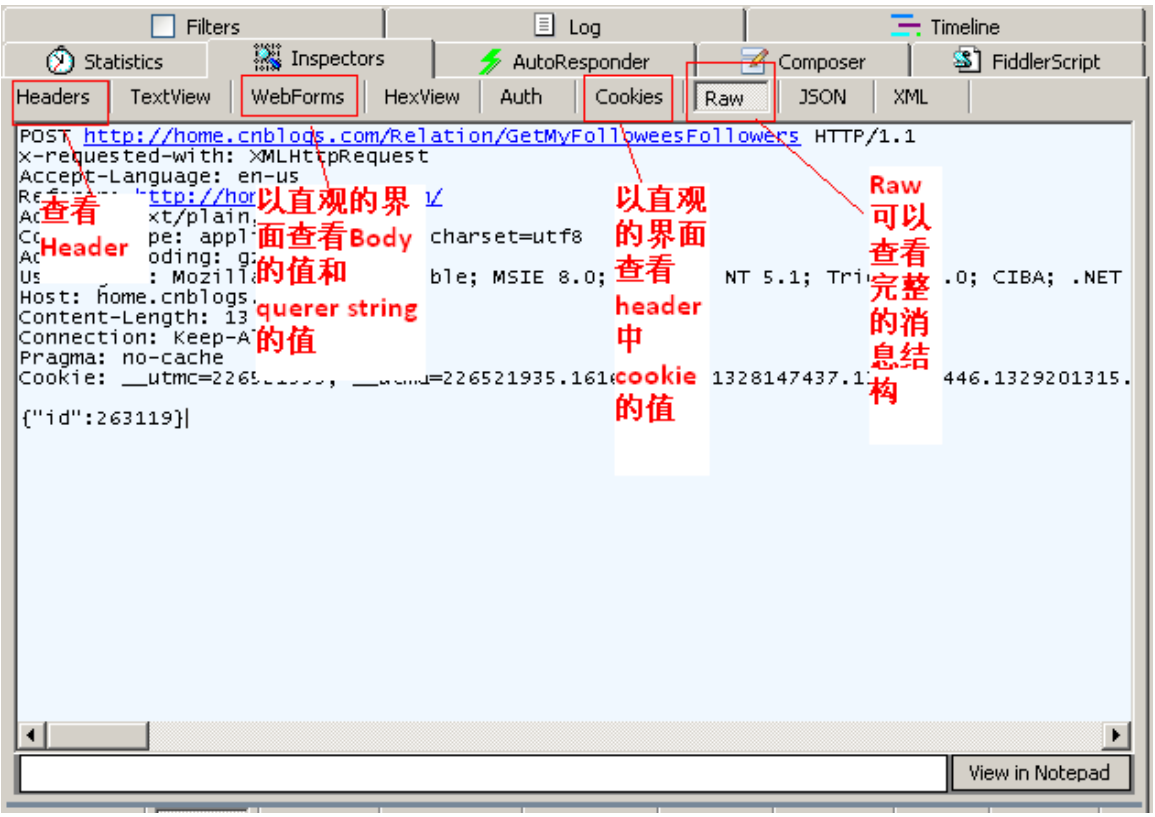


Fiddler 的基本界面

看看 Fiddler 的基本界面



Inspectors tab 下有很多查看 Request 或者 Response 的消息。 其中 Raw Tab 可以查看完整的消息，Headers tab 只查看消息中的 header.



请求（Request）部分详解

- Headers —— 显示客户端发送到服务器的 HTTP 请求的 header，显示为一个分级视图，包含了 Web 客户端信息、Cookie、传输状态等。
- Textview —— 显示 POST 请求的 body 部分，为文本。
- WebForms —— 显示请求的 GET 参数 和 POST body 内容。
- HexView —— 用十六进制数据显示请求。
- Auth —— 显示响应 header 中的 Proxy-Authorization(代理身份验证）和 Authorization(授权) 信息.
- Raw —— 将整个请求显示为纯文本。
- JSON - 显示 JSON 格式文件。
- XML —— 如果请求的 body 是 XML 格式，就是用分级的 XML 树来显示它。

Fiddler 图标含义

	正在将请求数据发往服务器
	正在从服务器下载返回数据
	请求过程中暂停
	返回过程中暂停
	请求中使用了 HTTP HEAD 方法；返回中应该没有 body 内容
	请求中使用了 HTTP CONNECT 方法，建立 HTTPS 连接通道
	返回的内容类型是 HTML
	返回的内容类型是图片
	返回的内容类型是 Javascript
	返回的内容类型是 CSS
	返回的内容类型是 XML
	普通的成功的返回
	返回内容为 HTTP/300,301,302,303 or 307 跳转
	返回内容为 HTTP/304：使用本地缓存

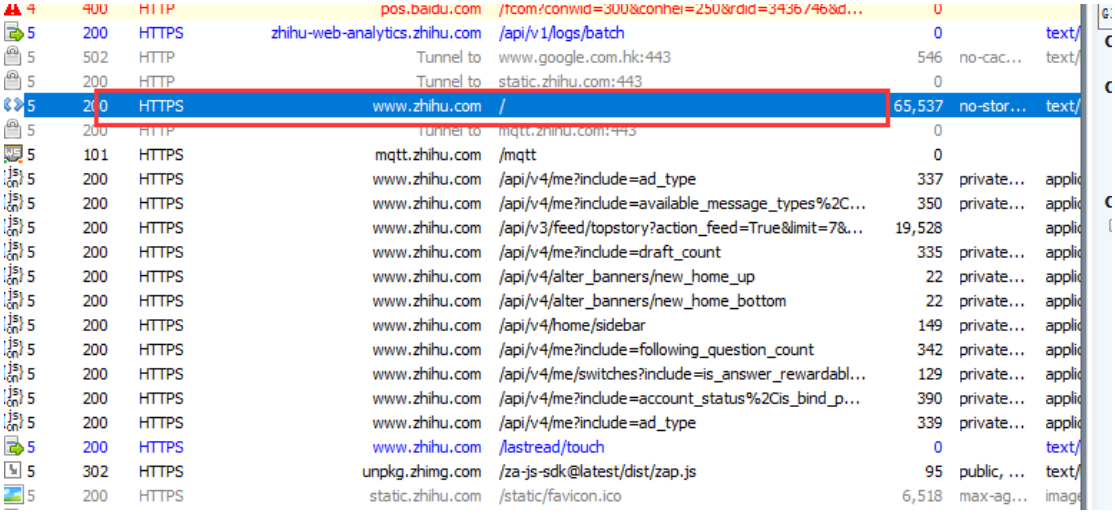
	返回内容为一个证书请求
	返回内容是服务器错误
	请求被客户端、Fiddler 或服务器中断

应用：抓包知乎网

浏览器中输入：https://www.zhihu.com/

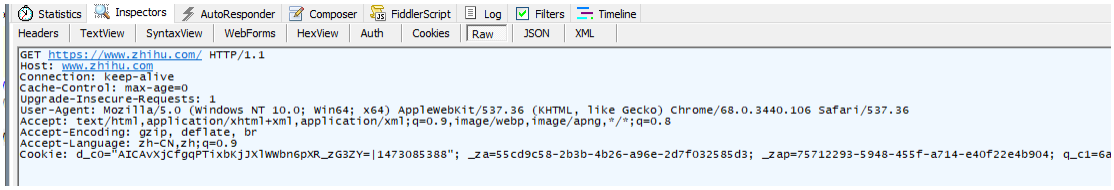
登陆

fiddler 中抓包



4	400	HTTP	pos.baidu.com	/fcom?conwid=300&conhei=250&rdid=3436746&d...	0		
5	200	HTTPS	zhihu-web-analytics.zhihu.com	/api/v1/logs/batch	0		text/
5	502	HTTP	Tunnel to	www.google.com.hk:443	546	no-cac...	text/
5	200	HTTP	Tunnel to	static.zhihu.com:443	0		
5	200	HTTPS	www.zhihu.com	/	65,537	no-stor...	text/
5	200	HTTP	Tunnel to	mqtt.zhihu.com:443	0		
5	101	HTTPS	mqtt.zhihu.com	/mqtt	0		
5	200	HTTPS	www.zhihu.com	/api/v4/me?include=ad_type	337	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/me?include=available_message_types%2C...	350	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v3/feed/topstory?action_feed=True&limit=7&...	19,528		applic
5	200	HTTPS	www.zhihu.com	/api/v4/me?include=draft_count	335	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/alter_banners/new_home_up	22	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/alter_banners/new_home_bottom	22	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/home/sidebar	149	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/me?include=following_question_count	342	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/me/switches?include=is_answer_rewardabl...	129	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/me?include=account_status%2Cis_bind_p...	390	private...	applic
5	200	HTTPS	www.zhihu.com	/api/v4/me?include=ad_type	339	private...	applic
5	200	HTTPS	www.zhihu.com	/lastread/touch	0		text/
5	302	HTTPS	unpkg.zhimg.com	/za-js-sdk@latest/dist/zap.js	95	public, ...	text/
5	200	HTTPS	static.zhihu.com	/static/favicon.ico	6,518	max-ag...	image

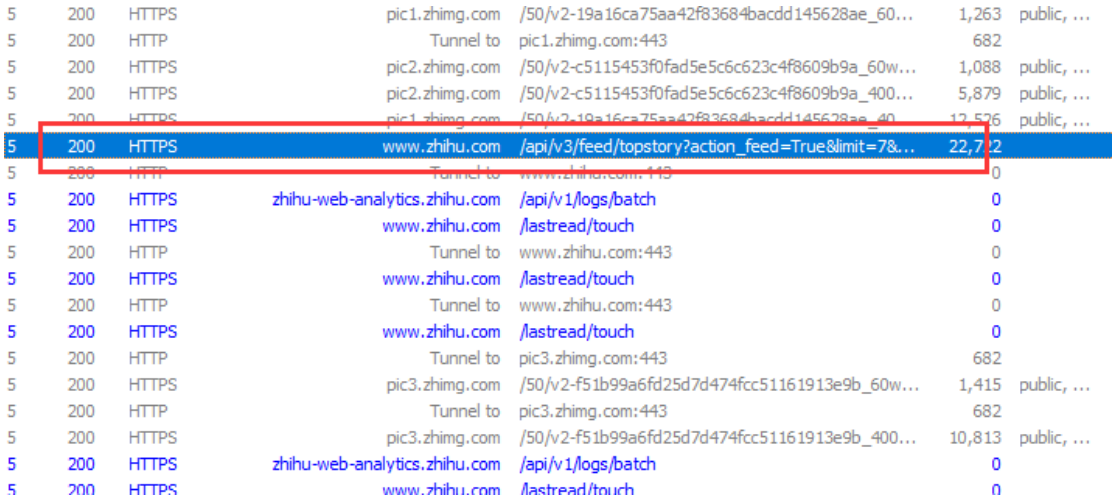
http request：



```
GET https://www.zhihu.com/ HTTP/1.1
Host: www.zhihu.com
Connection: keep-alive
Cache-Control: max-age=0
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/68.0.3440.106 Safari/537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8
Accept-Encoding: gzip, deflate, br
Accept-Language: zh-CN,zh;q=0.9
Cookie: d_c0="AICAvXjCfgqPTixbKjJXlWWbn6pXR_zG3ZY=|1473085388"; _za=55cd9c58-2b3b-4b26-a96e-2d7f032585d3; _zap=75712293-5948-455f-a714-e40f22e4b904; q_c1=6a
```

http response:

向下滚动页面，抓取 url：



5	200	HTTPS	pic1.zhimg.com	/50/v2-19a16ca75aa42f83684bacdd145628ae_60...	1,263	public, ...
5	200	HTTP	Tunnel to	pic1.zhimg.com:443	682	
5	200	HTTPS	pic2.zhimg.com	/50/v2-c5115453f0fad5e5c6c623c4f8609b9a_60w...	1,088	public, ...
5	200	HTTPS	pic2.zhimg.com	/50/v2-c5115453f0fad5e5c6c623c4f8609b9a_400...	5,879	public, ...
5	200	HTTPS	pic1.zhimg.com	/50/v2-19a16ca75aa42f83684bacdd145628ae_40...	12,526	public, ...
5	200	HTTPS	www.zhihu.com	/api/v3/feed/topstory?action_feed=True&limit=7&...	22,722	
5	200	HTTP	Tunnel to	www.zhihu.com:443	0	
5	200	HTTPS	zhihu-web-analytics.zhihu.com	/api/v1/logs/batch	0	
5	200	HTTPS	www.zhihu.com	/lastread/touch	0	
5	200	HTTP	Tunnel to	www.zhihu.com:443	0	
5	200	HTTPS	www.zhihu.com	/lastread/touch	0	
5	200	HTTP	Tunnel to	www.zhihu.com:443	0	
5	200	HTTPS	www.zhihu.com	/lastread/touch	0	
5	200	HTTP	Tunnel to	pic3.zhimg.com:443	682	
5	200	HTTPS	pic3.zhimg.com	/50/v2-f51b99a6fd25d7d474fcc51161913e9b_60w...	1,415	public, ...
5	200	HTTP	Tunnel to	pic3.zhimg.com:443	682	
5	200	HTTPS	pic3.zhimg.com	/50/v2-f51b99a6fd25d7d474fcc51161913e9b_400...	10,813	public, ...
5	200	HTTPS	zhihu-web-analytics.zhihu.com	/api/v1/logs/batch	0	
5	200	HTTPS	www.zhihu.com	/lastread/touch	0	

url 为：

https://www.zhihu.com/api/v3/feed/topstory?action_feed=True&limit=7&session_token=79ba91fe71c5cc32c8c6efb91ad303be&action=down&after_id=6&desktop=true

http response：

继续向下滚动，加载新的数据，抓取 url

url 是:

https://www.zhihu.com/api/v3/feed/topstory?action_feed=True&limit=7&session_token=79ba91fe71c5cc3207782e1807fd0299&action=down&after_id=13&desktop=true

http response:

3、结论

网页使用了 ajax 技术，通过接送数据进行更新

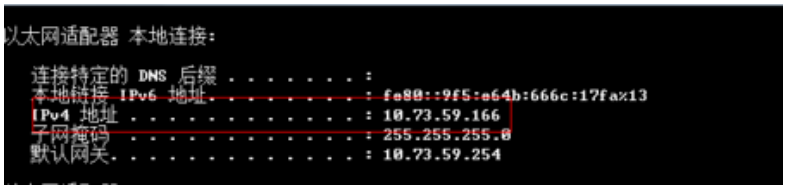
url 的区别：在于参数 after_id，每次加载 7 条数据，参数加 7

4、爬虫的实现

略

方式一、Fiddler 手机抓包配置

在命令提示符下输入 ipconfig 查看本机 IP:



打开 Android 设备的“设置”->“WLAN”，找到你要连接的网络，在上面长按，然后选择“修改网络”，弹出网络设置对话框，然后勾选“显示高级选项”。

在“代理”后面的输入框选择“手动”，在“代理服务器主机名”后面的输入框输入电脑的 ip 地址，在“代理服务器端口”后面的输入框输入 8888，然后点击“保存”按钮。

启动 Android 设备中的浏览器，访问网页即可在 Fiddler 中可以看到完成的请求和响应数据。

用 Fiddler 对 iPhone 手机应用进行抓包

基本流程差不多，只是手机设置不太一样：

iPhone 手机：点击设置 > 无线局域网 > 无线网络 > HTTP 代理 > 手动：

代理地址(电脑 IP)：192.168.xx.xxx　端口号：8888

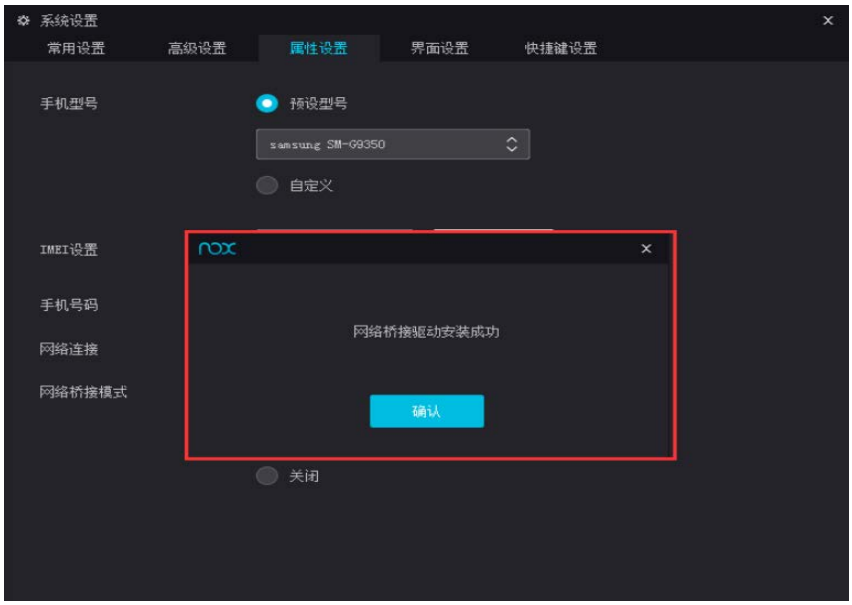
方式二、夜神模拟器安装配置

第一步:下载安装

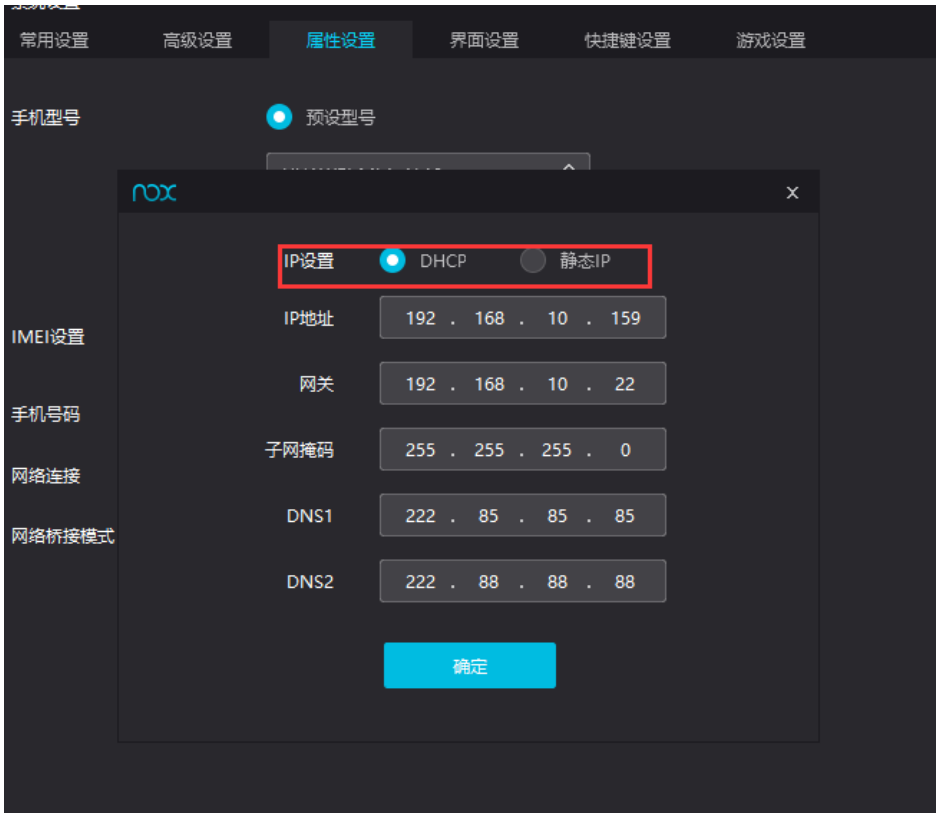
夜神模拟器下载完成之后，傻瓜式的安装一下！

第二步:配置桥接 实现互通

首先将当前手机网络桥接到本电脑网络 实现互通



配置 IP 地址，要配成和本机互通的网段



配置完成后打开主机 cmd 终端 ping 通 ok

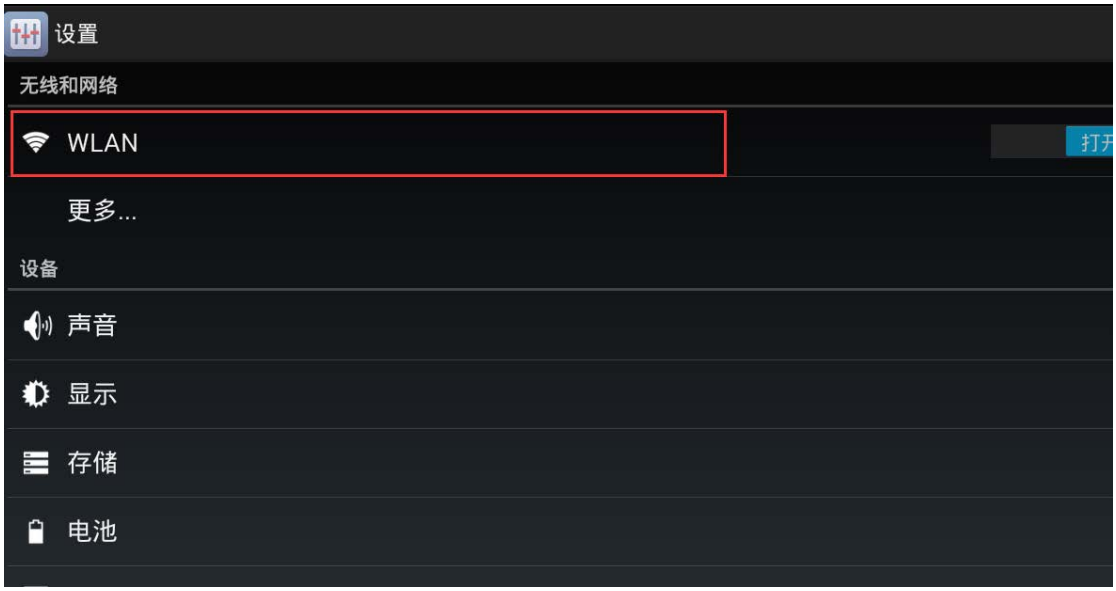
```
C:\Users\admin>ping 192.168.10.159

正在 Ping 192.168.10.159 具有 32 字节的数据:
来自 192.168.10.159 的回复: 字节=32 时间=1ms TTL=129
来自 192.168.10.159 的回复: 字节=32 时间<1ms TTL=129
来自 192.168.10.159 的回复: 字节=32 时间<1ms TTL=129
来自 192.168.10.159 的回复: 字节=32 时间<1ms TTL=129

192.168.10.159 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 0ms, 最长 = 1ms, 平均 = 0ms

C:\Users\admin>
```

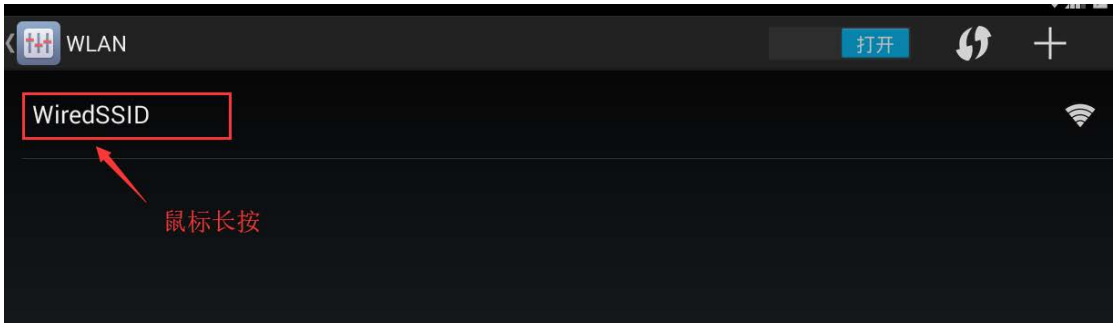
第三步:配置代理

输入 ipconfig 查看本机 IP

```
C:\Users\admin>ipconfig

Windows IP 配置


以太网适配器 本地连接:

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

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 . . . . . . . :

无线局域网适配器 无线网络连接:

   连接特定的 DNS 后缀 . . . . . . . : qikuedu.com
   本地链接 IPv6 地址. . . . . . . . : fe80::ac88:7912:6041:1b9%24
   IPv4 地址 . . . . . . . . . . . . : 192.168.10.103
   子网掩码  . . . . . . . . . . . . : 255.255.255.0
   默认网关. . . . . . . . . . . . . : 192.168.10.22
```

进入夜神模拟器–打开设置–打开 WLAN


设置
无线和网络
WLAN
打开
更多...
设备
声音
显示
存储
电池


WLAN
打开
WiredSSID
鼠标长按

WiredSSID
取消保存网络
修改网络

手机安装 Fiddler 的安全证书

使用 Android 手机的浏览器打开：http://xxx.xxx.xx.xx:8888（注：IP 跟端口都是你主机对应的）， 点"FiddlerRoot certificate" 然后安装证书

应用：爬取王者荣耀盒子 app

一、王者荣耀热点游戏新闻提取

打开手机 app 王者荣耀盒子，显示热点游戏新闻

Fiddler 手机抓包测试

WinConfig Replay Go Stream Decode Keep: All sessions Any Process Find Save Browse

#	Result	Protocol	Host	URL	Body	Caching	Content-Type	Process
7	200	HTTP	Tunnel to	ulogs.umeng.com:443	0			
8	200	HTTP	Tunnel to	ulogs.umengcloud.com:443	0			
9	200	HTTP	gamehelper.gm825.com	/apk/report/checkupdate?channel_id=90001a&app...	60	no-store, n...	application/...	
10	200	HTTP	gamehelper.gm825.com	/wzry/hot/information?channel_id=90001a&app_id...	6,904	no-store, n...	application/...	
11	200	HTTP	pic.wankacn.com	/2018-04-02_5ac1c7f64dca1.png	10,357	Expires: Su...	image/png	
12	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb0f9a024f.jpeg/thumblist	41,787	Expires: Fri...	image/jpeg	
13	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb1229969f.jpeg/thumblist	24,887	Expires: Fri...	image/jpeg	
14	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb0cb44c62.png/thumblist	48,199	Expires: Fri...	image/png	
15	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb0a13fb65.jpeg/thumblist	45,041	Expires: Fri...	image/jpeg	
16	502	HTTP	Tunnel to	stats.umsns.com:443	546	no-cache, ...	text/html; c...	
17	502	HTTP	Tunnel to	stats.umsns.com:443	546	no-cache, ...	text/html; c...	
18	200	HTTP	sspapi.gm825.com	/v2.0/getad?reqjson=Vy2PzAR6amD2TQjN1EY2%...	210		application/...	
19	200	HTTP	gamehelper.gm825.com	/wzry/article/50428.html?channel_id=90001a&app...	3,329	no-store, n...	text/html; c...	
20	200	HTTP	gamehelper.gm825.com	/static/wzry/css/article_detail.css?v=1.3	1,762	max-age=...	text/css	
21	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb1229969f.jpeg	83,601	no-cache	image/jpeg	
22	200	HTTP	gamehelper.gm825.com	/static/wzry/js/jquery-1.9.1.min.js	36,810	max-age=...	application/...	
23	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb122e9690.jpeg	99,860	no-cache	image/jpeg	
24	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb1233fbf6.jpeg	66,441	no-cache	image/jpeg	
25	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb12386dd4.jpeg	44,209	no-cache	image/jpeg	
26	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb123c07e3.jpeg	54,899	no-cache	image/jpeg	
27	200	HTTP	pic.wankacn.com	/2018-09-05_5b8fb12404bdc.jpeg	168,137	no-cache	image/jpeg	
28	200	HTTP	pic.wankacn.com	/2018-08-31_5b89010d5afa9.jpeg	54,453	no-cache	image/jpeg	

QueryString

Name	Value
channel_id	90001a
app_id	h9044j
game_id	7622
game_name	王者荣耀
vcode	13.0.1.0
version_code	13010
cuid	E223701EE539A9C66F8F69D3D934A169
ovr	4.4.2
device	HUAWEI _HUAWEI MLA-AL10
net_type	1
client_id	EyG/vecFAHM57Vi1on0eZQ==
info_ms	fH4d5UmMLAFOcyDSq1xHfw==
info_ma	ujY9SL1RpMSLiUJnkNXf3Vq6wHqeDWjzyY6dpsLRmgU=
mno	0
info_la	wPkXvFxUIsYfBxKnPvpGQg==
info_ci	wPkXvFxUIsYfBxKnPvpGQg==
mcc	0
clientversion	13.0.1.0

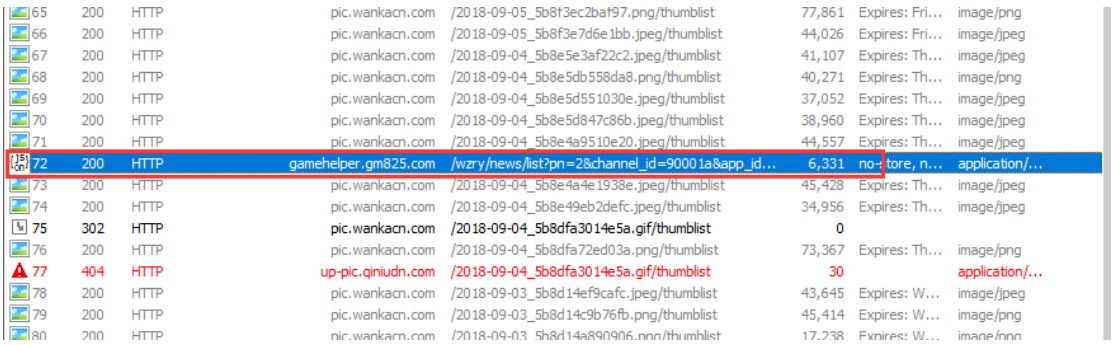
Transformer Headers TextView SyntaxView ImageView HexView WebView Auth Caching Cookies Raw JSON XML

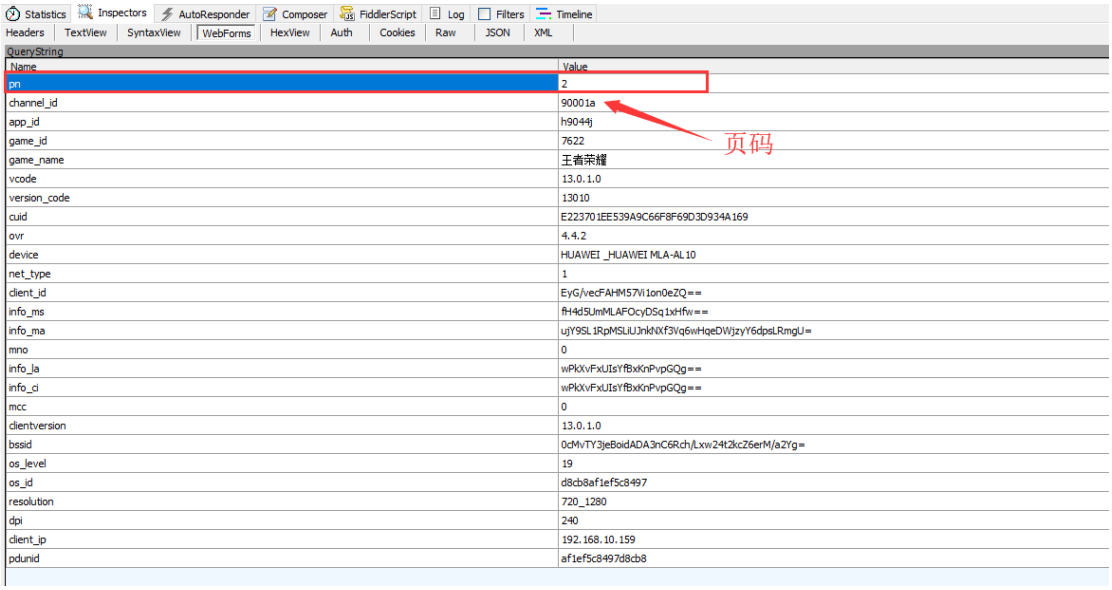

```
banner_list
error_code=0
first_sec_list
has_more=1
news_list
{}
cover=http://pic.wankacn.com/2018-09-04_5b8e5bf63a642.jpeg/thumblist
ctime=1536056160
description=王者荣耀盾山在体验服中的强度存在很大争议，盾山面对很多英雄都会处于弱势，但对某一类英雄却是天生克克，导致这个英雄在5V5实战中，特别影响体验感。不
detail_id=50407
is_top=1
keyword_id=2953
keyword_name=体验服更新
module_type=1
show_tags_type=keyword
show_type=1
source=互联网
source_type=2
tags_id=7
tags_name=体验服爆料
thumb_img
title=9月4日体验服更新：盾山大幅增强，最强辅助周五登场
video_url=
{}
{}
{}
```

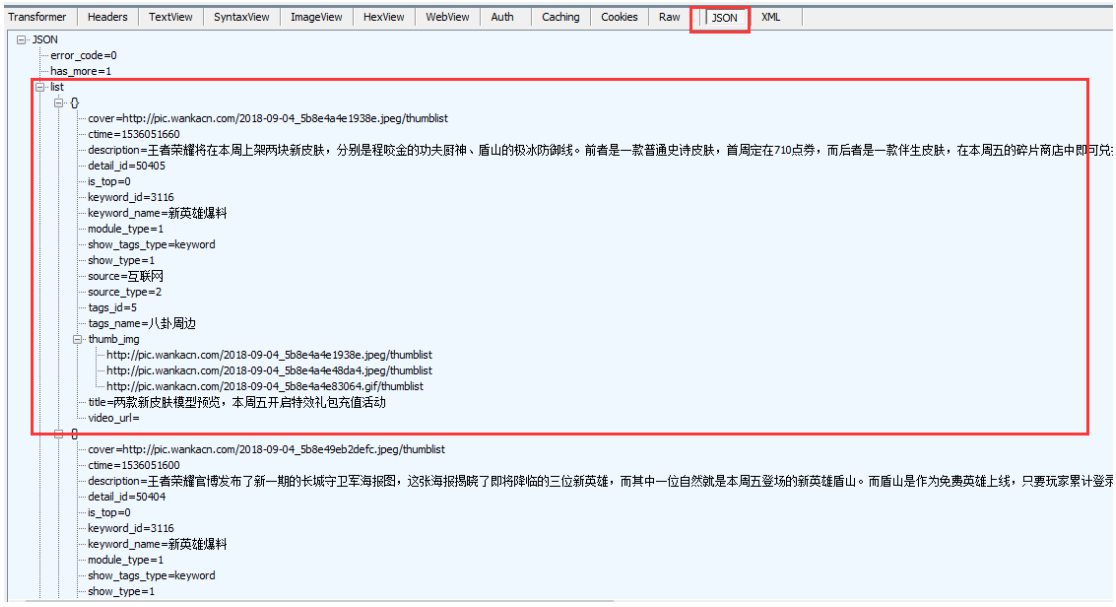

其 url 为:

http://gamehelper.gm825.com/wzry/hot/information?channel_id=90001a&app_id=h9044j&game_id=7622&game_name=%E7%8E%8B%E8%80%85%E8%8D%A3%E8%80%80&vcode=13.0.1.0&version_code=13010&cuid=E223701EE539A9C66F8F69D3D934A169&ovr=4.4.2&device=HUAWEI+_HUAWEI+MLA-AL10&net_type=1&client_id=EyG%2FvecFAHM57Vi1on0eZQ%3D%3D&info_ms=fH4d5UmMLAFOcyDSq1xHfw%3D%3D&info_ma=ujY9SL1RpMSLiUJnkNXf3Vq6wHqeDWjzyY6dpsLRmgU%3D&mno=0&info_la=wPkXvFxUIsYfBxKnPvpGQg%3D%3D&info_ci=wPkXvFxUIsYfBxKnPvpGQg%3D%3D&mcc=0&clientversion=13.0.1.0&bssid=0cMvTY3jeBoidADA3nC6Rch%2FLxw24t2kcZ6erM%2Fa2Yg%3D&os_level=19&os_id=d8cb8af1ef5c8497&resolution=720_1280&dpi=240&client_ip=192.168.10.159&pdunid=af1ef5c8497d8cb8

向下滑动页面，加载新的新闻，fiddler 继续抓取







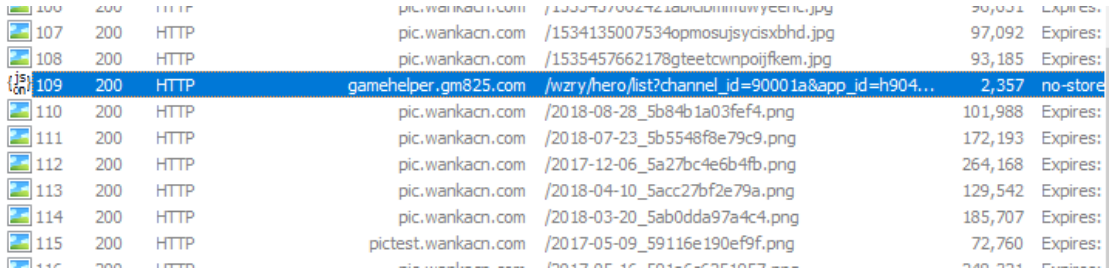
其 url：

http://gamehelper.gm825.com/wzry/news/list?pn=2&channel_id=90001a&app_id=h9044j&game_id=7622&game_name=%E7%8E%8B%E8%80%85%E8%8D%A3%E8%80%80&vcode=13.0.1.0&version_code=13010&cuid=E223701EE539A9C66F8F69D3D934A169&ovr=4.4.2&device=HUAWEI+_HUAWEI+MLA-AL10&net_type=1&client_id=EyG%2FvecFAHM57Vi1on0eZQ%3D%3D&info_ms=fH4d5UmMLAFOcyDSq1xHfw%3D%3D&info_ma=ujY9SL1RpMSLiUJnkNXf3Vq6wHqeDWjzyY6dpsLRmgU%3D&mno=0&info_la=wPkXvFxUIsYfBxKnPvpGQg%3D%3D&info_ci=wPkXvFxUIsYfBxKnPvpGQg%3D%3D&mcc=0&clientversion=13.0.1.0&bssid=0cMvTY3jeBoidADA3nC6Rch%2FLxw24t2kcZ6erM%2Fa2Yg%3D&os_level=19&os_id=d8cb8af1ef5c8497&resolution=720_1280&dpi=240&client_ip=192.168.10.159&pdunid=af1ef5c8497d8cb8

爬虫实现

略

二、王者荣耀英雄图片提取



107	200	HTTP	pic.wankacn.com	/1534135007534opmosujsycisxbhd.jpg	97,092	Expires:
108	200	HTTP	pic.wankacn.com	/1535457662178gteetcwnpoijfkem.jpg	93,185	Expires:
109	200	HTTP	gamehelper.gm825.com	/wzry/hero/list?channel_id=90001a&app_id=h904...	2,357	no-store
110	200	HTTP	pic.wankacn.com	/2018-08-28_5b84b1a03fef4.png	101,988	Expires:
111	200	HTTP	pic.wankacn.com	/2018-07-23_5b5548f8e79c9.png	172,193	Expires:
112	200	HTTP	pic.wankacn.com	/2017-12-06_5a27bc4e6b4fb.png	264,168	Expires:
113	200	HTTP	pic.wankacn.com	/2018-04-10_5acc27bf2e79a.png	129,542	Expires:
114	200	HTTP	pic.wankacn.com	/2018-03-20_5ab0dda97a4c4.png	185,707	Expires:
115	200	HTTP	pictest.wankacn.com	/2017-05-09_59116e190ef9f.png	72,760	Expires:

Statistics | Inspectors | AutoResponder | Composer | FiddlerScript | Log | Filters | Timeline

Headers | TextView | SyntaxView | WebForms | HexView | Auth | Cookies | Raw | JSON | XML

QueryString

Name	Value
channel_id	90001a
app_id	h9044j
game_id	7622
game_name	王者荣耀
vcode	13.0.1.0
version_code	13010
cuid	E223701EE539A9C66F8F69D3D934A169
ovr	4.4.2
device	HUAWEI _HUAWEI MLA-AL10
net_type	1
client_id	EyG/vecFAHM57Vi1on0eZQ==
info_ms	fH4d5UmMLAFOcyDSq1xHfw==
info_ma	ujY9SL1RpMSLiUJnkNXf3Vq6wHqeDWjzyY6dpsLRmgU=
mno	0
info_la	wPkXvFxUIsYfBxKnPvpGQg==
info_ci	wPkXvFxUIsYfBxKnPvpGQg==
mcc	0
clientversion	13.0.1.0
bssid	0cMvTY3jeBoidADA3nC6Rch/Lxw24t2kcZ6erM/a2Yg=
os_level	19
os_id	d8cb8af1ef5c8497
resolution	720_1280
dpi	240
client_ip	192.168.10.159
pdunid	af1ef5c8497d8cb8

Transformer | Headers | TextView | SyntaxView | ImageView | HexView | WebView | Auth | Caching | Cookies | Raw | JSON | XML


```
JSON
  error_code=0
  list
    {}
      cover=http://pic.wankacn.com/2018-08-28_5b84b1a03fef4.png
      hero_id=124
      name=盾山
      tags=2
      type
        3
        6
    {}
      cover=http://pic.wankacn.com/2018-07-23_5b5548f8e79c9.png
      hero_id=122
      name=司马懿
      tags=2
      type
        2
        4
    {}
      cover=http://pic.wankacn.com/2017-12-06_5a27bc4e6b4fb.png
      hero_id=112
      name=弈星
      tags=3
      type
        2
    {}
      cover=http://pic.wankacn.com/2018-04-10_5acc27bf2e79a.png
      hero_id=118
      name=米莱狄
      tags=3
      type
```


其 url：

http://gamehelper.gm825.com/wzry/hero/list?channel_id=90001a&app_id=h9044j&game_id=7622&game_name=%E7%8E%8B%E8%80%85%E8%8D%A3%E8%80%80&vcode=13.0.1.0&version_code=13010&cuid=E223701EE539A9C66F8F69D3D934A169&ovr=4.4.2&device=HUAWEI+_HUAWEI+MLA-AL10&net_type=1&client_id=EyG%2FvecFAHM57Vi1on0eZQ%3D%3D&info_ms=fH4d5UmMLAFOcyDSq1xHfw%3D%3D&info_ma=ujY9SL1RpMSLiUJnkNXf3Vq6wHqeDWjzyY6dpsLRmgU%3D&mno=0&info_la=wPkXvFxUIsYfBxKnPvpGQg%3D%3D&info_ci=wPkXvFxUIsYfBxKnPvpGQg%3D%3D&mcc=0&clientversion=13.0.1.0&bssid=0cMvTY3jeBoidADA3nC6Rch%2FLxw24t2kcZ6erM%2Fa2Yg%3D&os_level=19&os_id=d8cb8af1ef5c8497&resolution=720_1280&dpi=240&client_ip=192.168.10.159&pdunid=af1ef5c8497d8cb8

爬虫实现

略

Xposed 框架

起因

有些手机 APP 应用程序无法抓到数据包，因为 app 启用了 SSL Pinning（又叫“ssl 证书绑定“）

https 通讯过程

HTTPS 在建立 ssl 通道的过程中，当客户端向服务端发送了连接请求后，服务器会发送自己的证书(包括公钥、证书有效期、服务器信息等)给客户端，如果客户端是普通的浏览器，比如 IE 浏览器，则：

1. 使用内置的 CA 证书去校验服务器证书是否被信任，如果不被信任，则会弹出 https 的告警提示信息，由用户自己决定是否要继续.

2. 同样，用户也可以主动的将服务器证书导入到浏览器的受信任区，下次打开时该服务器证书将会自动被信任.

MITM 攻击

中间人攻击（Man-in-the-MiddleAttack，简称“MITM 攻击”）是一种“间接”的入侵攻击，通过拦截正常的网络通信数据，并进行数据篡改和嗅探，而通信的双方却毫不知情。如 SMB 会话劫持、DNS 欺骗等

伪造证书的中间人攻击可以劫持 https 原因正是因为上面的两点

SSLPinning

解决“中间人劫持+伪造证书”攻击的方法

客户端在收到服务器的证书后，对该证书进行强校验，验证该证书是不是客户端承认的证书，如果不是，则直接断开连接。同时使用足够的代码加密或混淆，这就是 SSLPinning

APP 是 HTTPS 的服务提供方自己开发的客户端，开发者可以先将自己服务器的证书打包内置到自己的 APP 中，或者将证书签名内置到 APP 中，当客户端在请求服务器建立连接期间收到服务器证书后，先使用内置的证书信息校验一下服务器证书是否合法，如果不合法，直接断开。

SSLPinning 的突破

使用 Xposed + JustTruestMe 来突破 SSL pinning

什么是 Xpose

框架是一款开源框架，其功能是可以在不修改 APK 的情况下影响程序运行(修改系统)的框架服务，基于它可以制作出许多功能强大的模块，且在功能不冲突的情况下同时运作。

基本上能够让我们修改整个设备系统的信息参数，如 display、board、dpi、厂商、gps 经纬度、手机 cpu 信息、局域网 ip 地址、系统版本号、操作系统信息、imsi 串号、mcc/mnc、手机型号、电话号码、手机 imei 串号、路由器 mac 地址、androidid、id、brand、网络类型、user、iso 国家码、指纹、软件版本、服务商信息、电话状态、系统版本名、运营商信息、设备信息、分辨率、wifi 名称、sn 序列号、硬件信息、sim 卡信息、等手机信息。

JustTruestMe

JustTrustMe 是一个用来禁用、绕过 SSL 证书检查的基于 Xposed 模块。JustTrustMe 是将 APK 中所有用于校验 SSL 证书的 API 都进行了 Hook 动态劫持方法，从而绕过证书检查。

安装

手机安装 xposed 框架需要 root 权限，**有一定的风险，手机有可能变砖头**，推荐使用模拟器，这里选择了夜神模拟器。

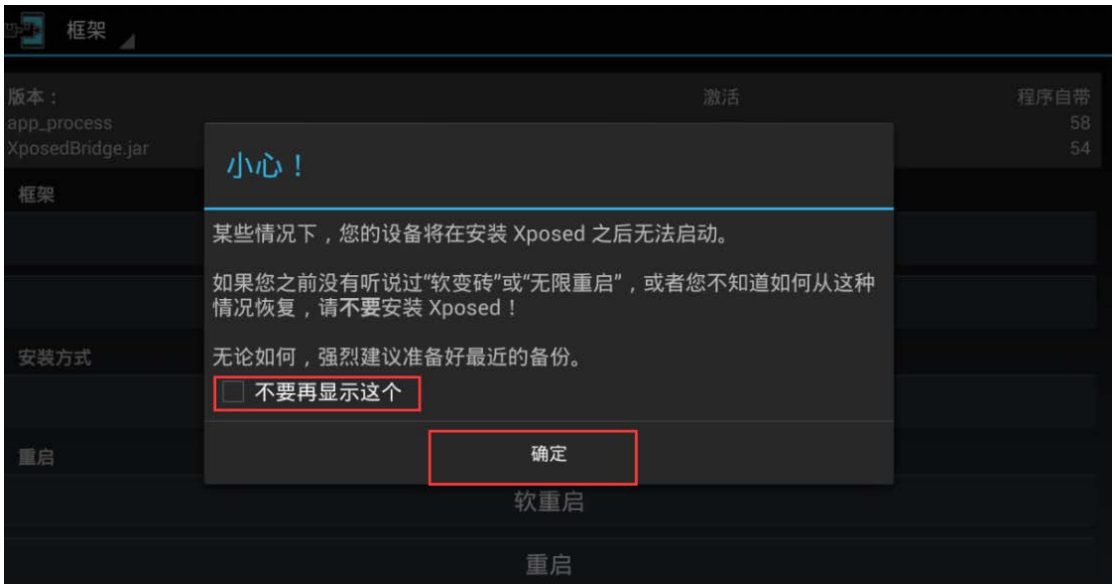
1、到官网下载 Xposed 框架安装器

2、打开夜神模拟器，把 XposedInstaller(完美适配夜神版).apk 拖进窗口进行安装

3、点击打开 xposed 安装器

4、选择框架

5、勾选‘不要再显示这个’项



6、选择安装更新

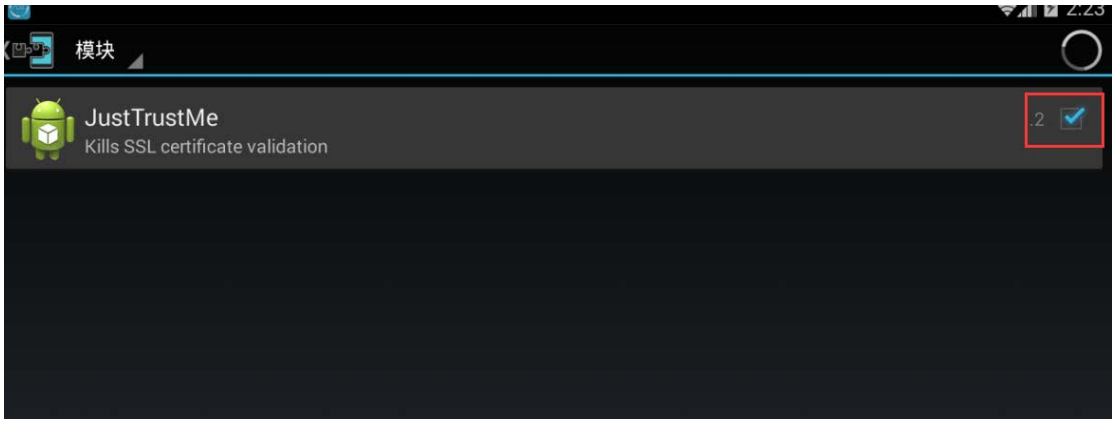
7、选择永久记住选择

8、安装

8、重启

9、向模拟器窗口中拖入 JustTrustMe.apk

10、打开安装器，进入模块，选择 JustTrustMe



案例：脉脉 app 职言爬取

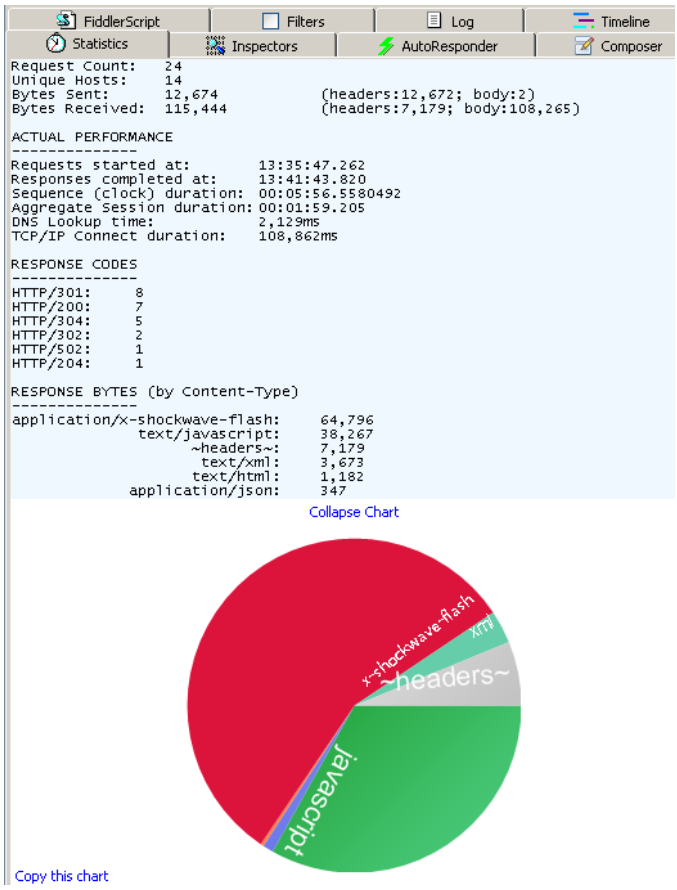
略

附录

Fiddler 的 HTTP 统计视图

通过陈列出所有的 HTTP 通信量，Fiddler 可以很容易的向您展示哪些文件生成了您当前请求的页面。使用 Statistics 页签，用户可以通过选择多个会话来得来这几个会话的总的信息统计，比如多个请求和传输的字节数。

选择第一个请求和最后一个请求，可获得整个页面加载所消耗的总体时间。从条形图表中还可以分别出哪些请求耗时最多，从而对页面的访问进行访问速度优化



QuickExec 命令行的使用

Fiddler 的左下角有一个命令行工具叫做 QuickExec,允许你直接输入命令。

常见得命令有

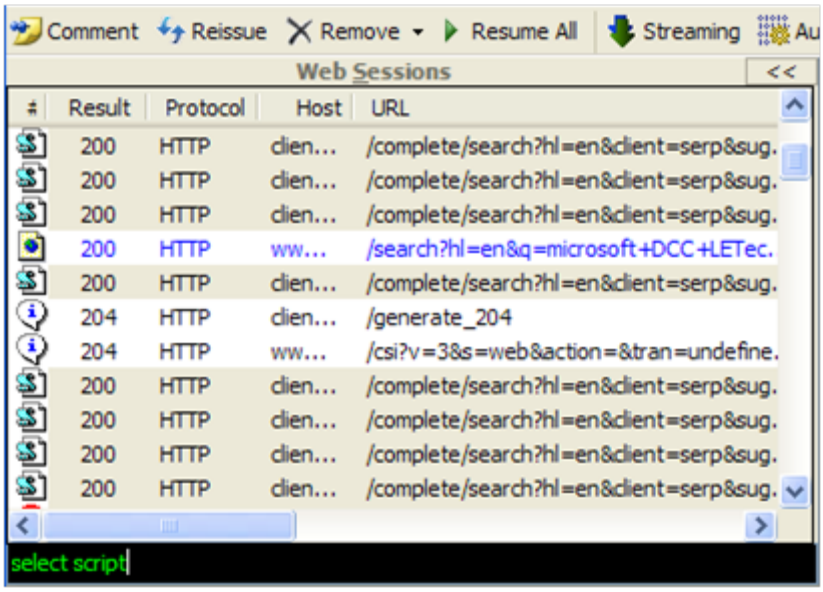
help 打开官方的使用页面介绍，所有的命令都会列出来

cls 清屏 (Ctrl+x 也可以清屏)

select 选择会话的命令

?.png 用来选择 png 后缀的图片

bpu 截获 request



Fiddler 中设置断点修改 Request

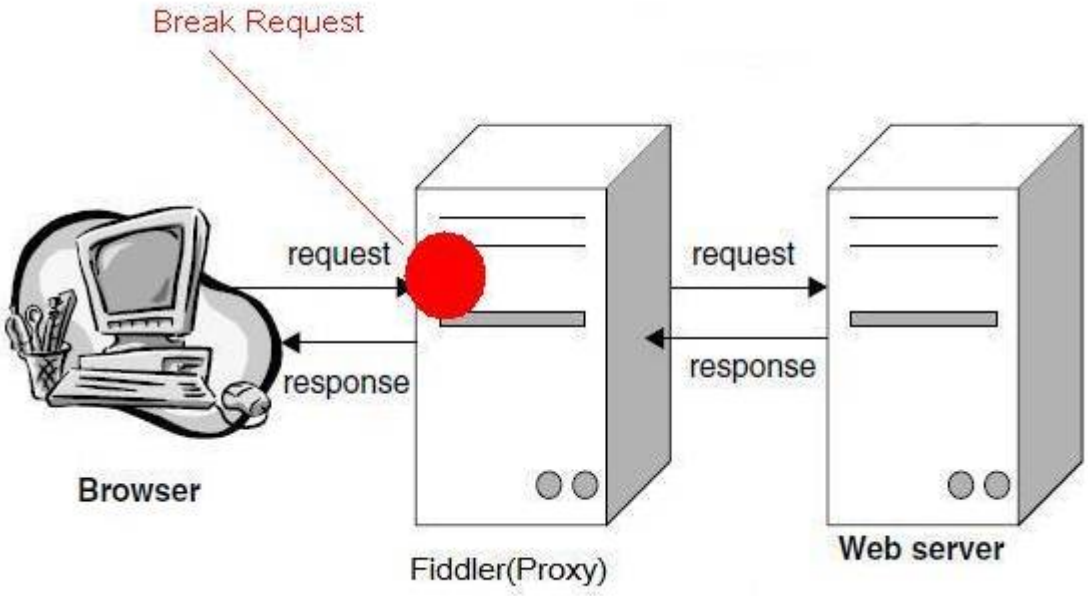
Fiddler 最强大的功能莫过于设置断点了，设置好断点后，你可以修改 httpRequest 的任何信息包括 host, cookie 或者表单中的数据。设置断点有两种方法

第一种：打开 Fiddler 点击 Rules-> Automatic Breakpoint ->Before Requests(这种方法会中断所有的会话)

如何消除命令呢？ 点击 Rules-> Automatic Breakpoint ->Disabled

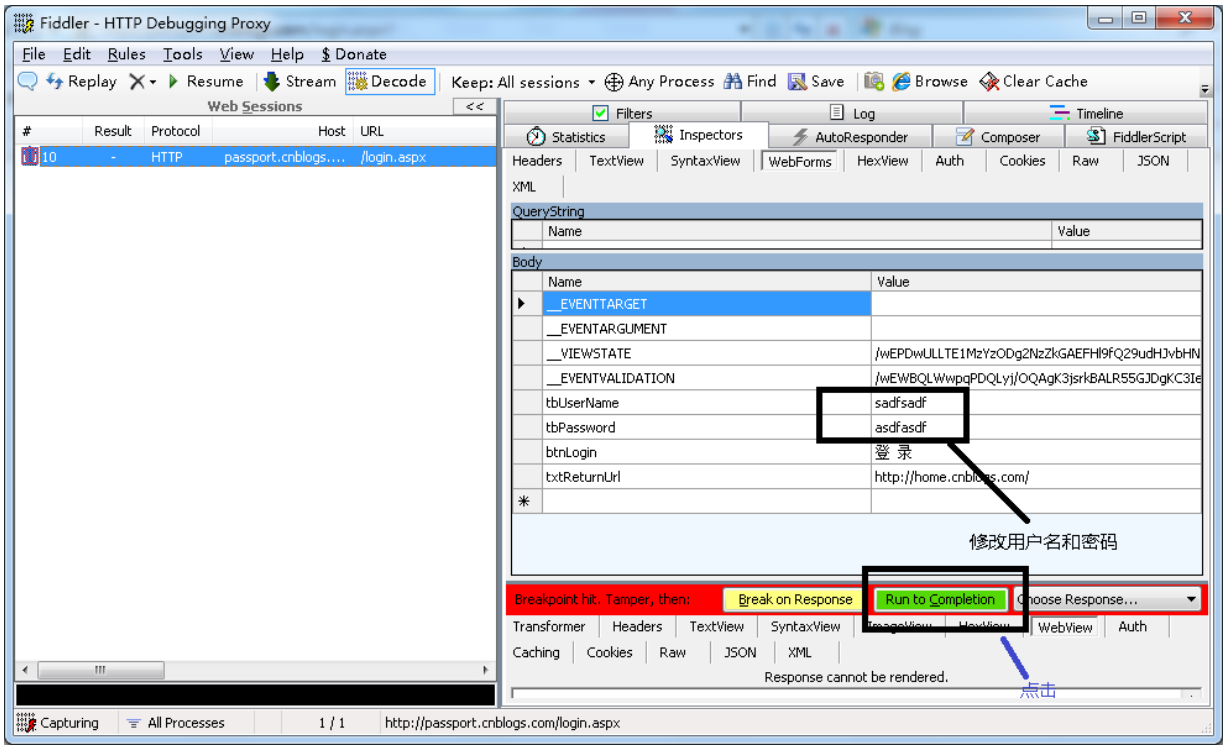
第二种：在命令行中输入命令：bpu www.baidu.com (这种方法只会中断 www.baidu.com)

如何消除命令呢？ 在命令行中输入命令 bpu



看个实例，模拟博客园的登录， 在 IE 中打开博客园的登录页面，输入错误的用户名和密码，用 Fiddler 中断会话，修改成正确的用户名密码。这样就能成功登录

1. 用 IE 打开博客园的登录界面 http://passport.cnblogs.com/login.aspx 2. 打开 Fiddler， 在命令行中输入 bpu http://passport.cnblogs.com/login.aspx 3. 输入错误的用户名和密码 点击登录 4. Fiddler 能中断这次会话，选择被中断的会话，点击 Inspectors tab 下的 WebForms tab 修改用户名密码，然后点击 Run to Completion 如下图所示。 5. 结果是正确地登录了博客园



Fiddler 中设置断点修改 Response

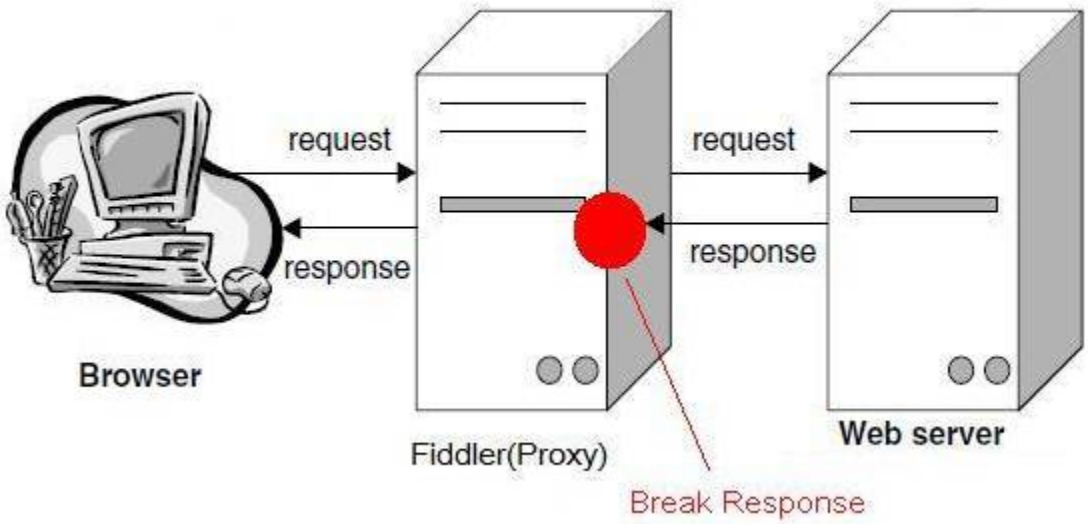
当然 Fiddler 中也能修改 Response

第一种：打开 Fiddler 点击 Rules-> Automatic Breakpoint ->After Response (这种方法会中断所有的会话)

如何消除命令呢？ 点击 Rules-> Automatic Breakpoint ->Disabled

第二种：在命令行中输入命令：bpafter www.baidu.com (这种方法只会中断 www.baidu.com)

如何消除命令呢？ 在命令行中输入命令 bpafter,



具体用法和上节差不多，就不多说了。

Fiddler 中创建 AutoResponder 规则

Fiddler 的 AutoResponder tab 允许你从本地返回文件，而不用将 http request 发送到服务器上。

看个实例. 1. 打开博客园首页，把博客园的 logo 图片保存到本地，并且对图片做些修改。

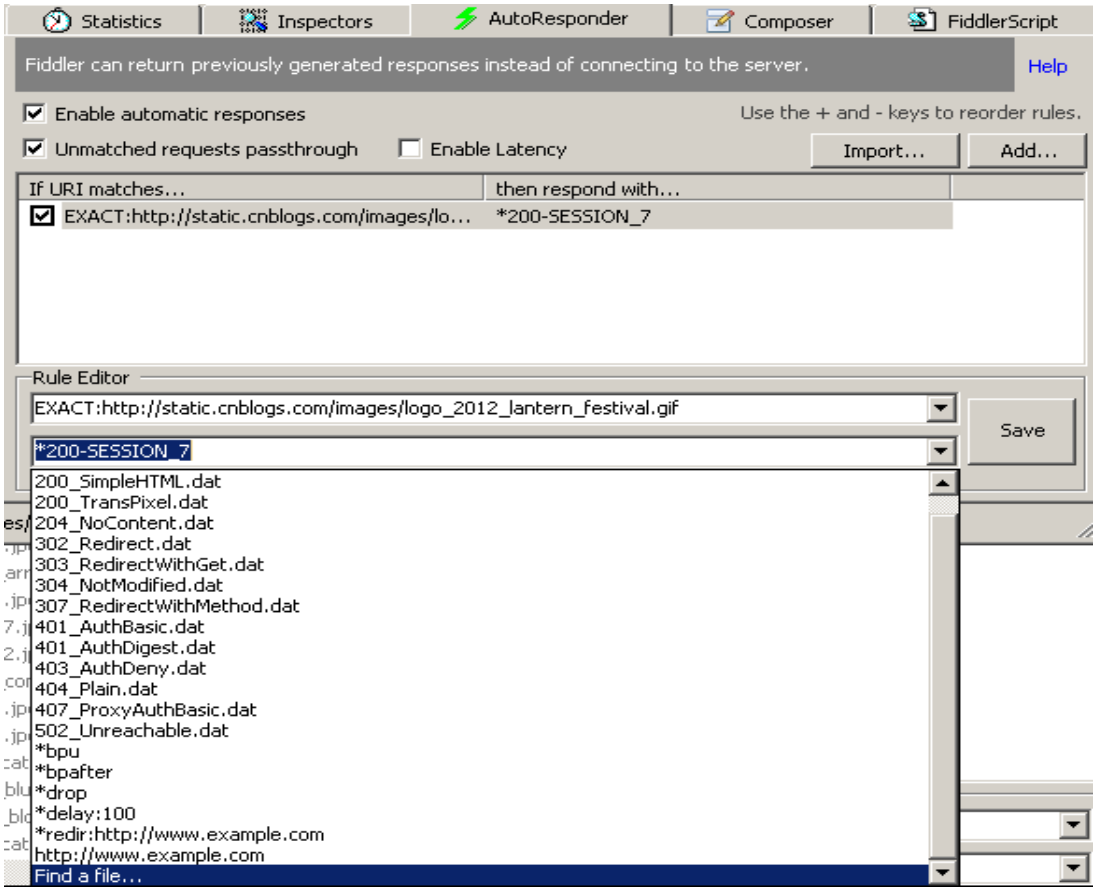
2. 打开 Fiddler 找到 logo 图片的会话，

http://static.cnblogs.com/images/logo_2012_lantern_festival.gif， 把这个会话拖到 AutoResponer Tab 下

3. 选择 Enable automatic reaponses 和 Unmatched requests passthrough

4. 在下面的 Rule Editor 下面选择 Find a file... 选择本地保存的图片. 最后点击 Save 保存下。

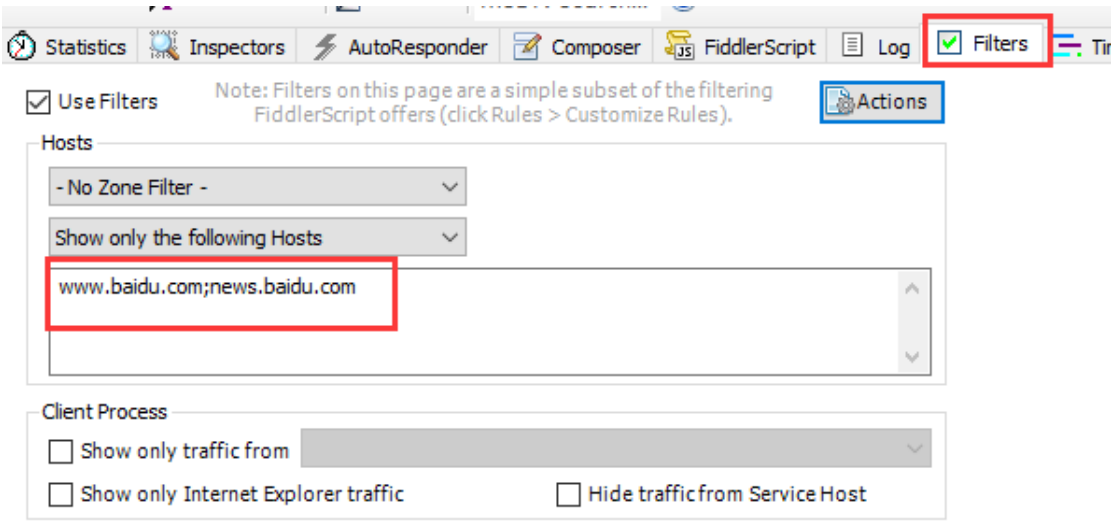
5. 再用 IE 博客园首页，你会看到首页的图片用的是本地的。


Statistics
Inspectors
AutoResponder
Composer
FiddlerScript
Fiddler can return previously generated responses instead of connecting to the server.
Help
Enable automatic responses
Use the + and - keys to reorder rules.
Unmatched requests passthrough
Enable Latency
Import...
Add...
If URI matches...
then respond with...
EXACT:http://static.cnblogs.com/images/lo...
*200-SESSION_7
Rule Editor
EXACT:http://static.cnblogs.com/images/logo_2012_lantern_festival.gif
Save
*200-SESSION_7
200_SimpleHTML.dat
200_TransPixel.dat
204_NoContent.dat
302_Redirect.dat
303_RedirectWithGet.dat
304_NotModified.dat
307_RedirectWithMethod.dat
401_AuthBasic.dat
401_AuthDigest.dat
403_AuthDeny.dat
404_Plain.dat
407_ProxyAuthBasic.dat
502_Unreachable.dat
*bpu
*bpafter
*drop
*delay:100
*redir:http://www.example.com
http://www.example.com
Find a file...

博客园 - 程序员的网上家园 - Windows Internet Explorer
http://www.cnblogs.com/
Bing
File Edit View Favorites Tools Help
Favorites
字符集...
Ent...
Ash...
AM ...
(原创...
CLR...
Cod...
Goo...
在线翻...
博客园 - 程序员的网上家园
Page Safety Tools
代码改变世界
[登录 · 注册]
博客园
游澳洲，赢大奖 和全球青年一决高下
Imagine Cup 2012 "创新杯" 微软全球学生大赛
园子 新闻 博问 闪存 网摘 招聘 专题 知识库
网站分类
.NET技术(2)
编程语言(5)
软件设计(0)
Web前端(9)
企业信息化(3)
手机开发(6)
软件工程(1)
首页 精华区 候选区 新闻 关注
RSS
【编辑推荐】一个老程序员的十年回顾
(3277) »
[推荐头条]深入理解JavaScript系列(18)：面向对象编程之ECMAScr...(8/473,10) »
[评论头条]观察 HTML中id和name 的差异，被微软忽悠过的同学自动举手(51/2673,4)
找找看
Google Custom Search
搜索
移动开发专区：iOS Android WP7
年终职场大关怀，邀请好友送礼品
新版博问上线，帮您解决IT难题
2012年MPD软件工作坊北京站开启
最新新闻
Done
Internet
100%

Fiddler 中如何过滤会话

每次使用 Fiddler，打开一个网站，都能在 Fiddler 中看到几十个会话，看得眼花缭乱。最好的办法是过滤掉一些会话，比如过滤掉图片的会话. Fiddler 中有过滤的功能



在 Filers 面板中勾选“Use Filters”，并在 Hosts 区域，设置以下三个选项：

1).第一项有三个选项，分别是：

“No zone filter”

“Show Only Intranet Hosts”

“Show Only Internet Hosts”，不做更改；

2).第二项有四个选项，分别是：

“No Host Filter”不设置 hosts 过滤；

“Hide The Following Hosts”隐藏过滤到的域名；

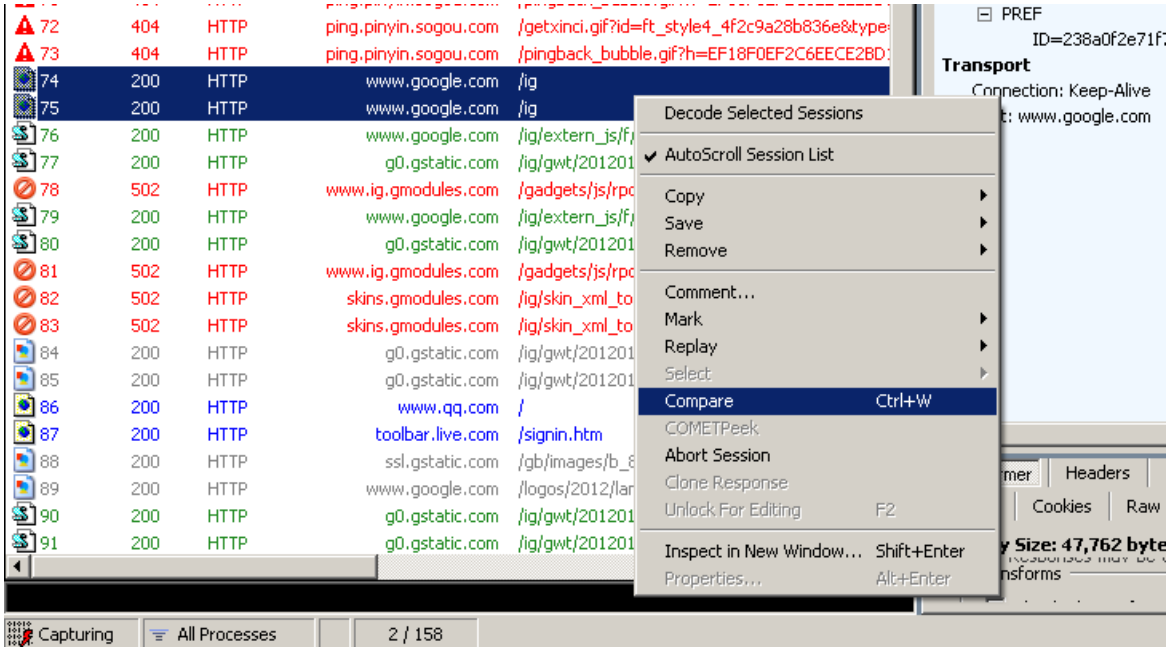
“Show Only The Following Hosts”只显示过滤到的域名；

“Flag The Following Hosts”标记过滤到的域名；

选中“Show Only The Following Hosts”，在文本框内输入需要过滤的域名，多个域名使用分号分割。fiddler 默认会检查 http 头中设置的 host，强制显示 http 地址中的域名。

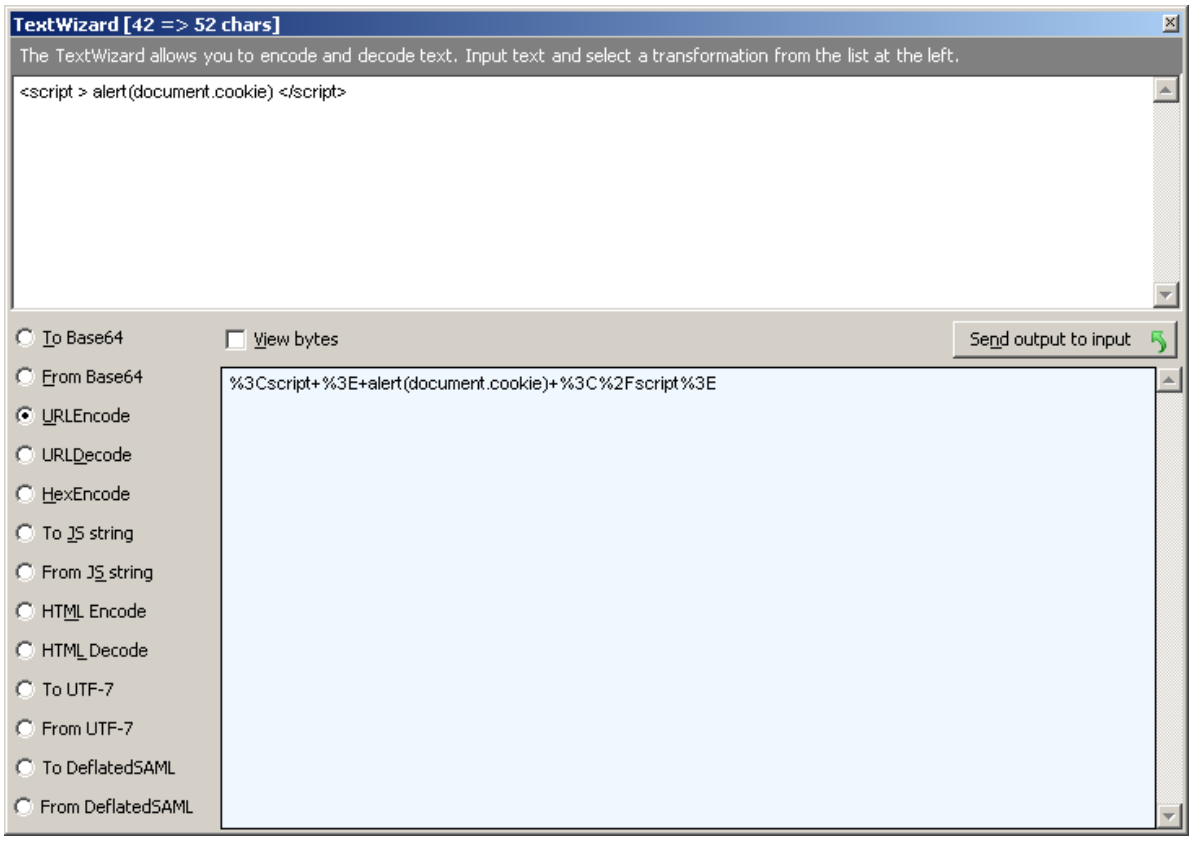
Fiddler 中会话比较功能

选中 2 个会话，右键然后点击 Compare，就可以用 WinDiff 来比较两个会话的不同了（当然需要你安装 WinDiff）



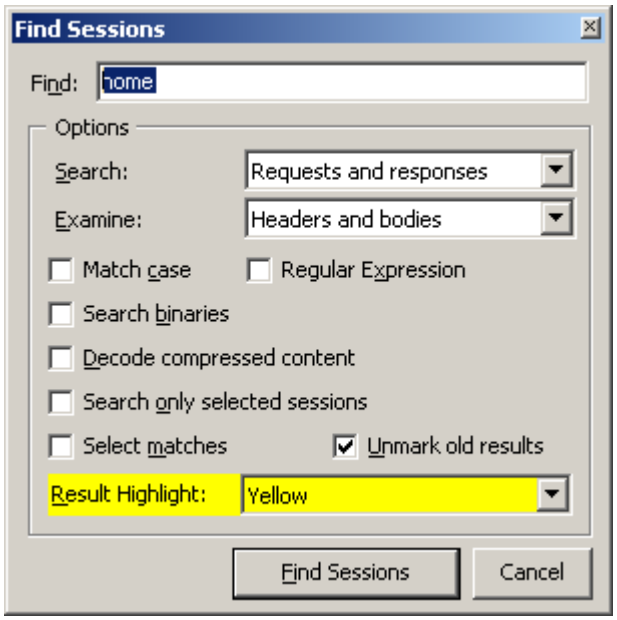
Fiddler 中提供的编码小工具

点击 Fiddler 工具栏上的 TextWizard，这个工具可以 Encode 和 Decode string.



Fiddler 中查询会话

用快捷键 Ctrl+F 打开 Find Sessions 的对话框，输入关键字查询你要的会话。 查询到的会话会用黄色显示



Fiddler 中保存会话

有些时候我们需要把会话保存下来，以便发给别人或者以后去分析。 保存会话的步骤如下：

选择你想保存的会话，然后点击 File->Save->Selected Sessions

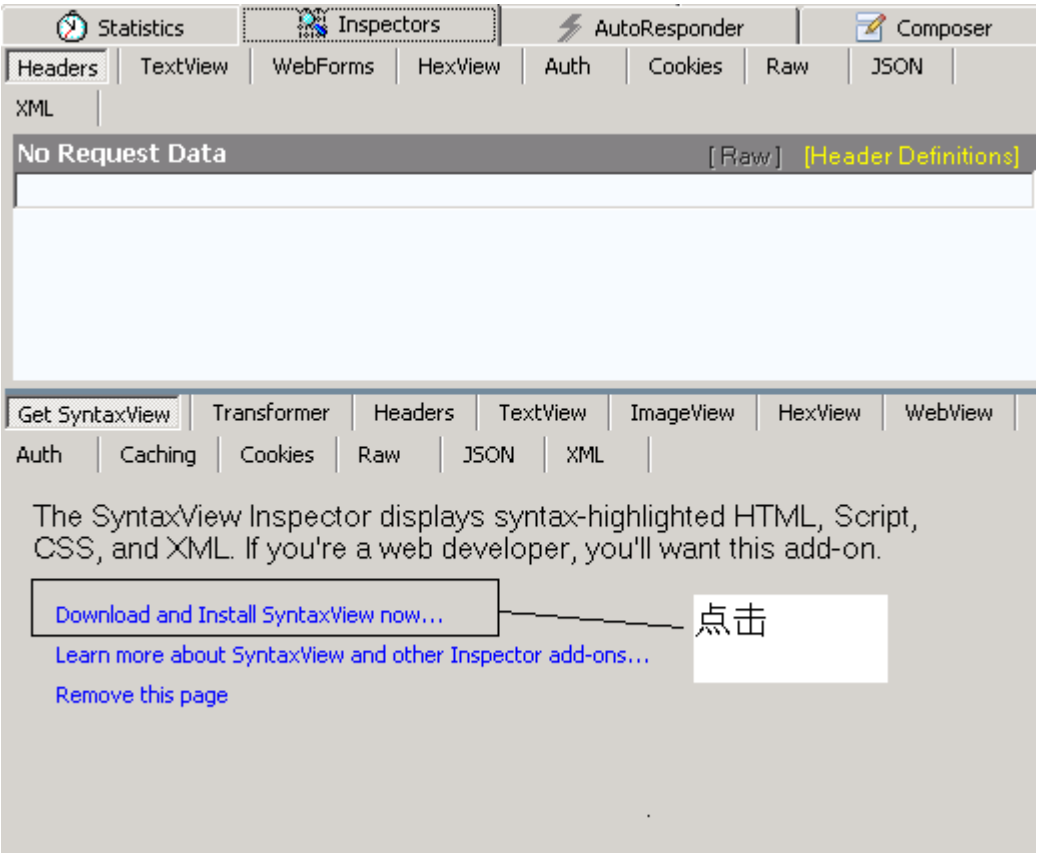
Fiddler 的 script 系统

Fiddler 最复杂的莫过于 script 系统了 官方的帮助文

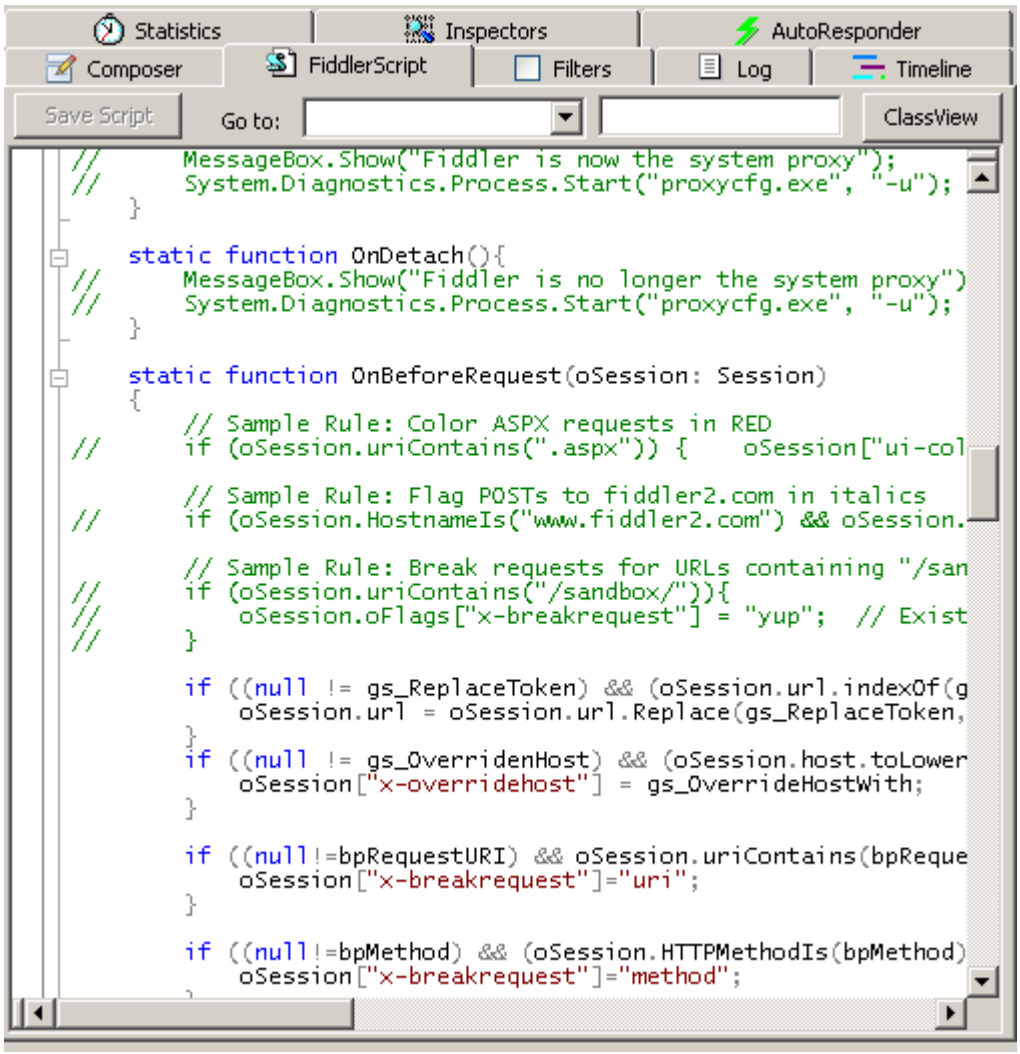
档: http://www.fiddler2.com/Fiddler/dev/ScriptSamples.asp

首先先安装 SyntaxView 插件，Inspectors tab->Get SyntaxView tab->Download and

Install SyntaxView Now... 如下图



安装成功后 Fiddler 就会多了一个 Fiddler Script tab，如下图



在里面我们就可以编写脚本了， 看个实例 让所有 cnblogs 的会话都显示红色。

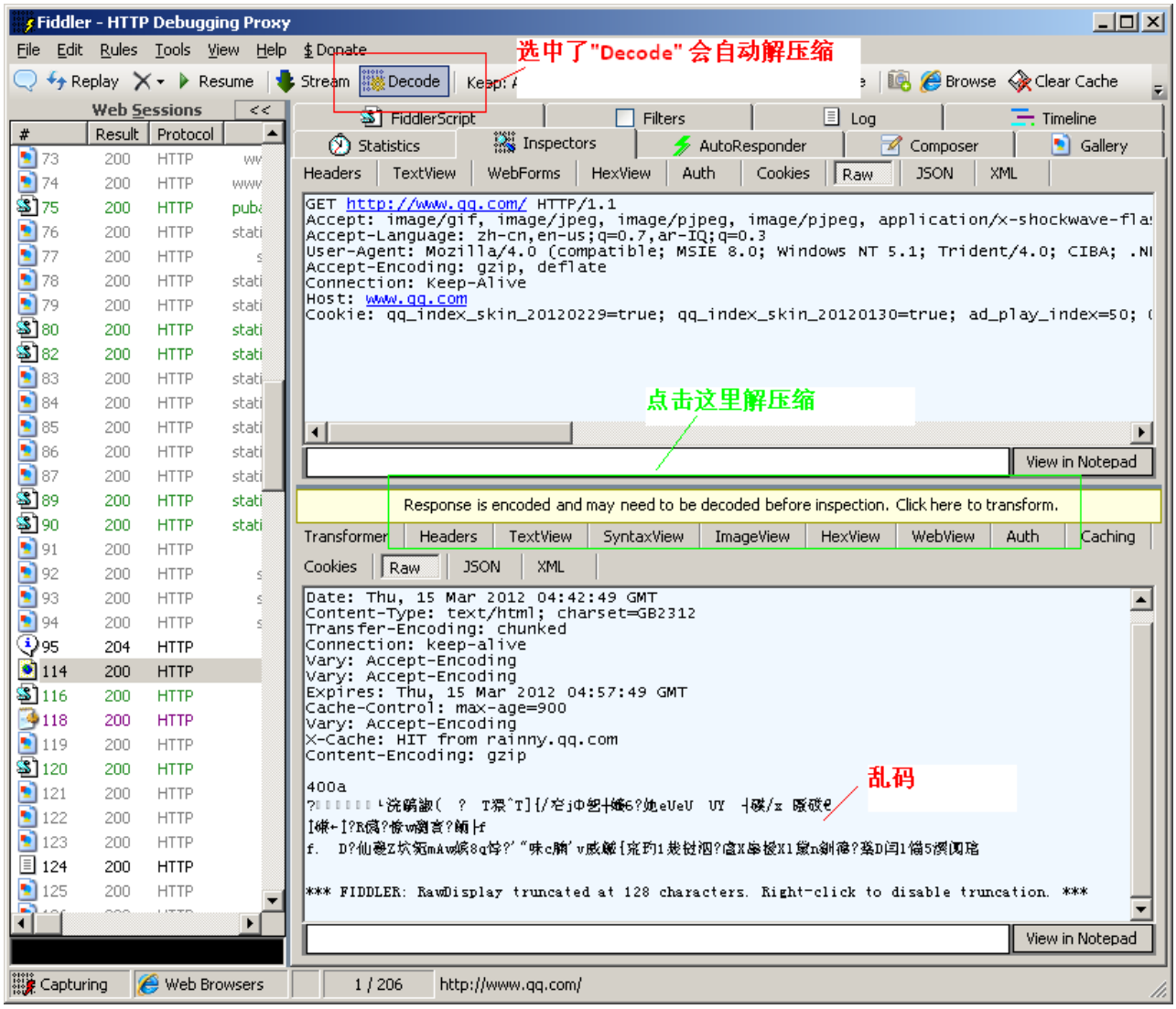
把这段脚本放在 OnBeforeRequest(oSession: Session) 方法下，并且点击"Save script"

```
    if (oSession.HostnameIs("www.cnblogs.com"))
{                oSession["ui-color"] = "red";          }
```

这样所有的 cnblogs 的会话都会显示红色

Response 是乱码的

有时候我们看到 Response 中的 HTML 是乱码的， 这是因为 HTML 被压缩了， 我们可以通过两种方法去解压缩。

1. 点击 Response Raw 上方的"Response is encoded any may need to be decoded before inspection. click here to transform"

2. 选中工具栏中的"Decode"。 这样会自动解压缩。