1 / 验证码（最简单有效的防护），采用点触验证，滑动验证或第三方验证码服务，普通验证码很容易被破解

2 / 频率，限制同设备，同IP等发送次数，单点时间范围可请求时长

3 / 归属地，检测IP所在地是否与手机号归属地匹配；IP所在地是否是为常在地

4 / 可疑用户，对于可疑用户要求其主动发短信（或其他主动行为）来验证身份

5 / 黑名单，对于黑名单用户，限制其操作，API接口直接返回success，1可以避免浪费资源，2混淆黑户判断

—

6 / 签名，API接口启用签名策略，签名可以保障请求URL的完整安全，签名匹配再继续下一步操作

7 / token，对于重要的API接口，生成token值，做验证

—

8 /  https，启用https，https 需要秘钥交换，可以在一定程度上鉴别是否伪造IP

9 / 代码混淆，发布前端代码混淆过的包

10 /  风控，大量肉鸡来袭时只能受着，同样攻击者也会暴露意图，分析意图提取算法，分析判断是否为恶意 如果是则断掉；异常账号及时锁定；或从产品角度做出调整，及时止损。

11 / 数据安全，数据安全方面做策略，攻击者得不到有效数据，提高攻击者成本

12 / 恶意IP库，https://threatbook.cn/，过滤恶意IP

---------------------

限制爬虫可以通过以下几种方法实现：

　　1） 设置robots.txt文件

　　限制爬虫最简单的方法是设置robots.txt文件。robots.txt文件是搜索引擎爬虫访问网站的时候要查看的第一个文件，它告诉爬虫程序在服务器上什么文件是可以被查看的，如设置Disallow： /，则表示所有的路径均不能查看。遗憾的是并不是所有的搜索引擎爬虫会遵守这个规则，因此仅仅设置robots文件是不够的。

　　2） User Agent识别与限制

　　要对不理会robots.txt文件的爬虫访问进行限制，首先要把爬虫流量与普通用户的访问流量进行区分，即对其进行识别。一般的爬虫程序都可以通过其HTTP请求中的User Agent字段进行识别，该字段使服务器能够识别客户使用的操作系统及版本、CPU类型、浏览器及版本、浏览器渲染引擎、浏览器语言、浏览器插件等。爬虫的User Agent字段一般与浏览器的有所不同，如Google搜索引擎爬虫User Agent字段中会有类似Googlebot的字符串，如User-Agent： Googlebot/2.1 （ [http://www.google.com/bot.html](http://www.google.com/bot.html" \t "https://blog.csdn.net/wangkeke1996/article/details/_blank)），百度搜索引擎爬虫则会有类似Baiduspider的字符串。不少Web服务器软件如 Apache，可以设置通过User Agent字段进行访问过滤，可以比较有效的限制大部分爬虫的访问。

　　3） 通过访问行为特征识别与限制

　　对于在HTTP请求的User Agent字段刻意伪装成浏览器的爬虫，可以通过其访问行为特征进行识别。爬虫程序的访问一般是有规律性的频率比较高，区别于真实用户浏览时的随意性与低 频率。对这类爬虫的限制原理与DDoS攻击的防御原理很相似，都基于统计数据。对于这类爬虫的限制只能通过应用识别设备、IPS等能够做深度识别的网络设 备来实现。用[网络设备](http://network.chinaitlab.com/" \t "https://blog.csdn.net/wangkeke1996/article/details/_blank)限制网络爬虫，不仅比较全面，而且非常适合在多服务器情况下进行统一的管理，避免多服务器单独管理有可能造成的疏漏。

**Mozilla/5.0 (iPhone; CPU iPhone OS 11\_0 like Mac OS X) AppleWebKit/604.1.38 (KHTML, like Gecko) Version/11.0 Mobile/15A372 Safari/604.1**

**Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/73.0.3683.86 Safari/537.36**