阿里云
爬虫风险管理

快速入门

文档版本：20190124

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **禁止:** 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **警告:** 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 **注意:** 您也可以通过按**Ctrl** + **A**选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 **确定**。
`courier` 字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

1 步骤1：添加域名

购买并开通云盾爬虫风险管理（Anti-Bot Service，简称Anti-Bot）产品后，您需要将您的网站域名接入Anti-Bot实例实现恶意Bot流量防护。

> **注意:**
> 由于Anti-Bot与Web应用防火墙（WAF）采用同样的转发配置，因此如果您已经购买开通阿里云WAF产品，您无需为已经接入WAF防护的网站域名重新配置Anti-Bot接入，已接入WAF的网站域名将直接显示在防爬虫风险管理控制台中，且域名来源为**云端同步**。

准备工作

在将您的网站域名接入Anti-Bot实例前，您需要完成以下准备工作：

- 需要防护恶意Bot流量的网站域名。

  > **注意:**
  > 对于中国大陆地区的Anti-Bot实例，所添加的网站域名必须已经通过工信部备案（无需通过阿里云备案）；对于海外地区的Anti-Bot实例，对于网站域名没有备案要求。

- 该网站域名的源站服务器IP。

  > **注意:**
  > 对于一个网站域名，Anti-Bot支持配置多达20个源站IP。

- 对于支持HTTPS业务的网站，您还需要准备该域名绑定的证书和私钥信息。
- 该网站域名的DNS解析记录的编辑权限。即确保您在该域名所在的域名解析服务商（如万网、新网、DNSPod）提供的管理控制台中可以修改该域名的DNS解析记录。

操作步骤

1. 登录爬虫风险管理控制台，选择您的Anti-Bot实例所在的地区。
2. 定位到**域名接入**页面，单击**添加域名**。
3. 填写网站域名相关配置信息，单击**下一步**。

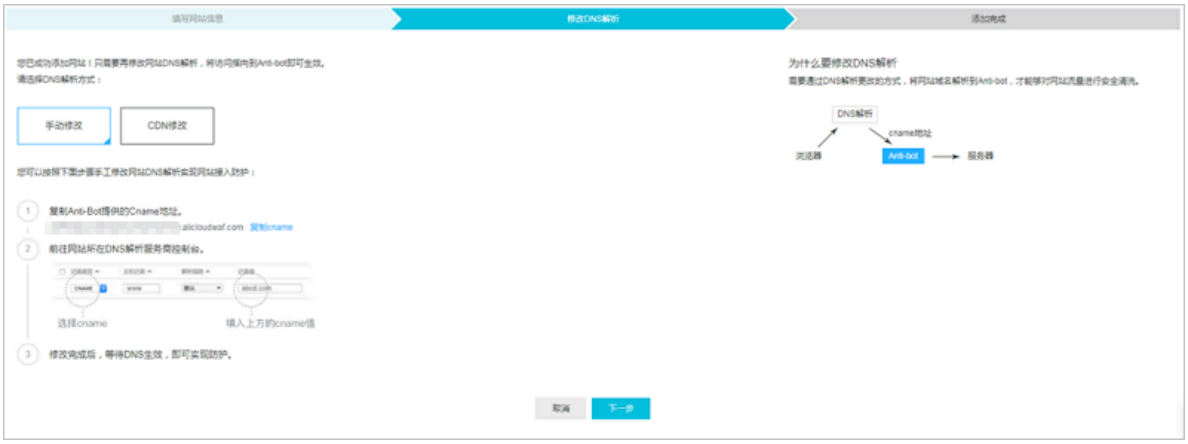
参数	描述	说明
域名	您想要防护的网站域名	支持填写泛域名（如*.aliyundemo.cn），Anti-Bot将自动匹配该泛域名的二级域名。 **注意:** 如果您同时配置泛域名和精确域名（如*.aliyundemo.cn和abc.aliyundemo.c Bot将优先使用精确域名所配置的转发规则和防护策略。
协议类型	该网站支持的协议类型。如果您的网站支持HTTPS加密认证，请勾选HTTPS协议，并在添加网站后上传证书和私钥。	如果勾选HTTPS协议类型，可使用高级设置实现HTTP强制跳转和HTTP回源等功能，保证访问平滑。
服务器地址	该网站域名的源站服务器地址。	网站接入Anti-Bot实例进行防护后，Anti-Bot会将过滤后的访问请求转发至该服务器地址。
	（推荐）勾选**IP**，填写源站服务器的IP（如云服务器ECS实例的IP、负载均衡SLB实例的IP等），配置成功后Anti-Bot将请求转发至该源站IP。	最多支持配置20个源站IP。如果配置多个回源IP，Anti-Bot将自动进行健康检查和负载均衡。
	勾选**其它地址**，填写服务器回源域名（如对象存储OSS的CNAME等），配置成功后Anti-Bot将请求转发至该域名。	服务器回源域名不应和要防护的网站域名相同。
服务器端口	该网站域名的源站端口。网站接入Anti-Bot实例进行防护后，Anti-Bot会将过滤后的访问请求转发至该端口。	HTTP协议默认为80端口；HTTPS协议默认为443端口。如果您想要使用其它端口，可单击**自定义**进行添加。

参数	描述	说明
Anti-Bot前是否有七层代理（高防/CDN等）	根据该网站业务的实际情况勾选。	如果该网站在Anti-Bot前需要配置其它七层代理进行转发，请务必勾选是，否则Anti-Bot将无法获取访问该网站的客户端真实IP信息。
负载均衡算法	如果配置多个源站IP，勾选**IP hash**或轮询，Anti-Bot将根据所选择的方式分发访问请求，实现负载均衡。	-

4. 选择修改该域名DNS解析的方式（手动修改、CDN修改），单击下一步，完成添加域名信息。

注意:

在修改域名的*DNS*解析前，建议您先在网站的源站服务器上配置放行*Anti-Bot*回源*IP*段并在本地环境中验证域名转发配置是否生效，确保该网站的访问请求可以正常转发回源站服务器。



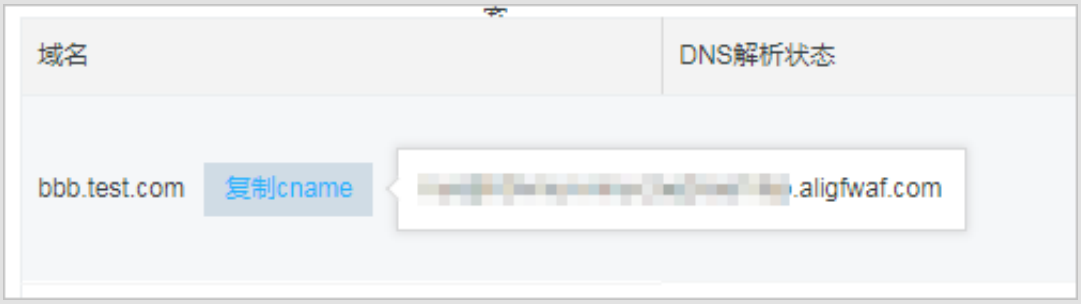
获取所分配的CNAME地址

网站域名信息添加成功后，Anti-Bot将自动为该域名分配一个CNAME。通过将该网站域名的访问解析到该CNAME，所有访问请求即可先经过该Anti-Bot实例后被转发至源站服务器，实现安全防护。

> **注意:**
> 如果想要获取该域名所分配到的Anti-Bot实例的IP，您通过Ping该域名所分配到CNAME解析得到的IP即是所分配的Anti-Bot实例的IP。一般情况下，所分配的Anti-Bot实例IP不会频繁变更。

登录爬虫风险管理控制台，在域名接入页面，选择已添加的网站域名，将鼠标移至**复制cname**按钮上即可查看Anti-Bot为该网站域名分配的CNAME。

> **注意:**
> 单击**复制cname**可将该CNAME复制到剪贴板中。



上传HTTPS证书和私钥（仅适用于HTTPS站点）

如果您添加的网站支持HTTPS加密认证，且已勾选HTTPS协议类型，在添加域名后，您还需要在爬虫风险管理控制台上传相应的证书和私钥，否则将无法通过HTTPS协议访问该网站。

域名信息添加成功后，在域名接入页面中，该网站域名的HTTPS协议状态将显示为异常，提示您当前证书配置有误。

参考以下步骤，为该网站域名上传证书和私钥：

1. 登录爬虫风险管理控制台，在域名接入页面，选择该网站域名。
2. 单击HTTPS协议状态右侧的上传证书按钮。
3. 在更新证书对话框中，选择上传方式。

   - 勾选手动上传，填写证书名称，将该网站域名所绑定的证书文件和私钥文件中的文本内容分别复制粘贴到证书文件和私钥文件文本框中。

     注意:

     对于一般格式的证书（如.pem、.cer、.crt等格式的证书），您可用文本编辑器直接打开证书文件复制其中的文本内容；对于其他格式的证书（如.pfx、.p7b等格式的证书）的证书，则需要将证书文件转换成.pem格式后再用文本编辑器打开来复制其中的文本内容。关于证书格式的转换方式，请查看*HTTPS*证书转换成*PEM*格式。

     注意:

     如果该HTTPS证书有多个证书文件（如证书链），需要将证书文件中的文本内容拼接合并后粘贴至证书文件文本框中。

     证书文件文本内容样例

     ```
     -----BEGIN CERTIFICATE-----
     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx8ixZJ4krc+1M+
     j2kcubVpsE2
     cgHdj4v8H6jUz9Ji4mr7vMNS6dXv8PUkl/qoDeNGCNdyTS5NIL5ir+g92cL8IGOk
     jgvhlqt9vc
     65Cgb4mL+n5+DV9uOyTZTW/MojmlgfUekC2xiXa54nxJf17Y1TADGSbyJbsC0Q9
     nIrHsPl8YKk
     vRWvIAqYxXZ7wRwWWmv4TMxFhWRiNY7yZIo2ZUhl02SIDNggIEeg==
     -----END CERTIFICATE-----
     ```

     私钥文件文本内容样例

     ```
     -----BEGIN RSA PRIVATE KEY-----
     DADTPZoOHd9WtZ3UKHJTRgNQmioPQn2bqdKHop+B/dn/4VZL7Jt8zSDGM9sTMThL
     yvsmLQKBgQ
     ```

```
Cr+ujntC1kN6pGBj2Fw2l/EA/W3rYEce2tyhjgmG7rZ+A/jVE9fld5sQra6ZdwBcQJ
aiygoIYo
aMF2EjRwc0qwHaluq0C15f6ujSoHh2e+D5zdmkTg/3NKNjqNv6xA2gYpinVDz
FdZ9Zujxvuh9o
4Vqf0YF8bv5UK5G04RtKadOw==
-----END RSA PRIVATE KEY-----
```

- 如果该网站域名所绑定的HTTPS证书已添加至该云账号的云盾证书服务进行管理，勾选**选择已有证书**，直接选择该证书即可。

4. 单击**保存**，该域名所绑定的证书和私钥即上传成功，HTTPS协议状态更新为**正常**。

2 步骤2：配置放行Anti-Bot回源IP段

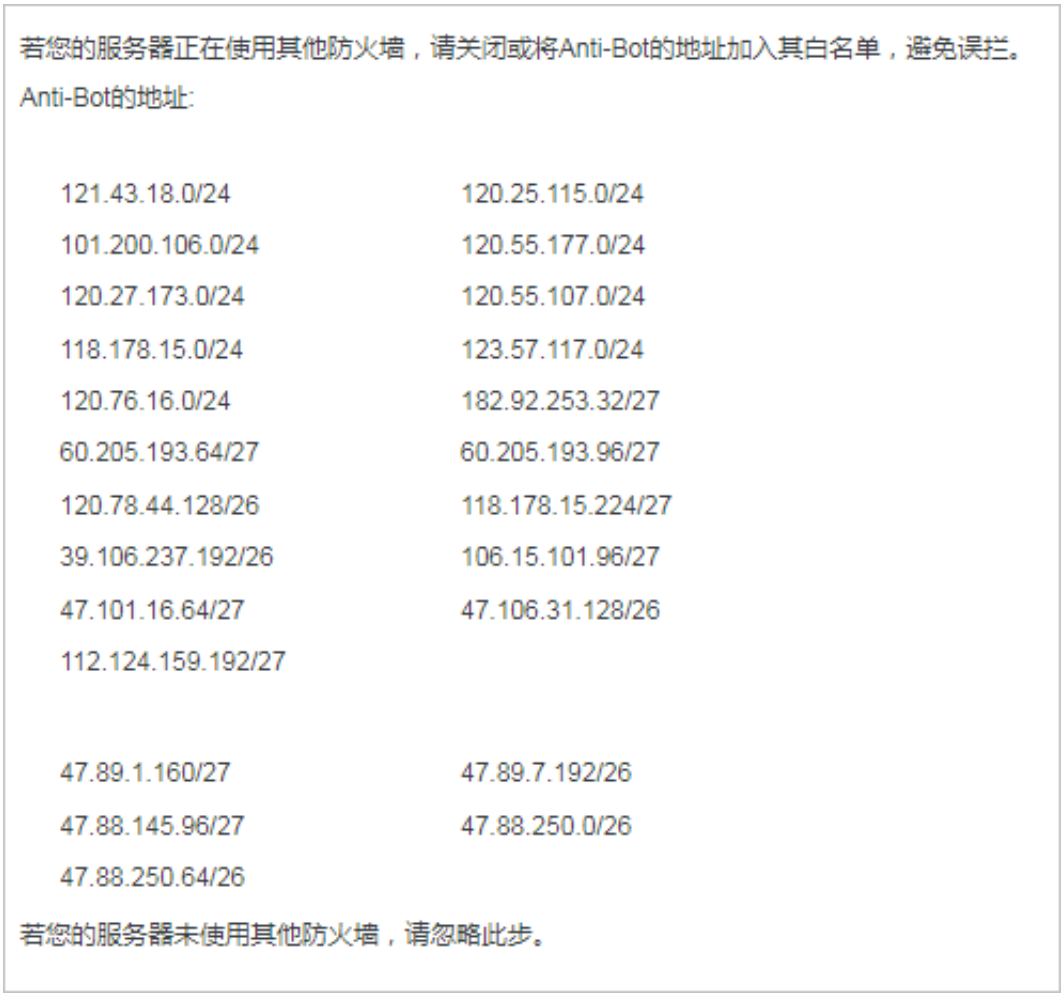
您的网站域名成功接入爬虫风险管理（Anti-Bot Service，简称Anti-Bot）防护后，访问您网站的所有请求将先流转到Anti-Bot实例，经Anti-Bot实例过滤恶意Bot流量后再返回到源站服务器。其中，流量经Anti-Bot实例返回源站的操作称为回源。

由于Anti-Bot实例的IP数量有限，源站服务器收到的所有请求都将来自这些IP，在源站服务器上的安全软件（如安全狗、云锁）看来，这种情况是很可疑的，且可能会屏蔽Anti-Bot实例的回源IP。因此，在接入Anti-Bot防护后，您需要在源站服务器的其它防火墙、安全软件上设置放行所有Anti-Bot回源IP。

操作步骤

爬风险管理控制台提供了最新的回源IP段列表，您可以参照以下步骤进行操作：

1. 在您将域名信息添加至Anti-Bot实例时，您会看到Anti-Bot实例的回源IP端。



2. 在源站服务器的防火墙、安全软件上，将步骤1中的IP段添加到白名单中。

常见问题

什么是回源IP？

回源IP是Anti-Bot用来代理客户端请求服务器时用的源IP，在服务器看来，接入Anti-Bot防护后所有源IP都会变成Anti-Bot实例的回源IP，而真实的客户端地址会被加在HTTP头部的XFF字段中。

在接入Anti-Bot实例后，您应确保源站已将Anti-Bot的全部回源IP放行（加入白名单），不然可能会出现网站打不开或打开极其缓慢等情况。

为何要放行回源IP段？

接入Anti-Bot实例后，Anti-Bot作为一个反向代理存在于客户端和服务器之间，服务器的真实IP被隐藏起来，客户端只能看到Anti-Bot实例，而看不到源站服务器。

因此，在源站（真实服务器）看来，所有的请求源IP都会变成Anti-Bot实例的回源IP段。

由于来源的IP变得更加集中，频率会变得更快，源站服务器上的防火墙或安全软件很容易认为这些IP在发起攻击，从而限制来自这些IP的访问。一旦Anti-Bot实例的回源IP被封禁，由Anti-Bot实例转发的请求将无法得到源站的正常响应，故务必确保Anti-Bot的回源IP在源站上不会被拦截。

3 步骤3：本地验证转发配置生效

在把业务流量切到爬虫风险管理（Anti-Bot Service，简称Anti-Bot）实例之前，建议您先通过本地验证确保一切配置正常，Anti-Bot转发正常。本地验证需要在本地模拟接入Anti-Bot，然后访问被防护网站，验证Anti-Bot实例正常转发访问请求。

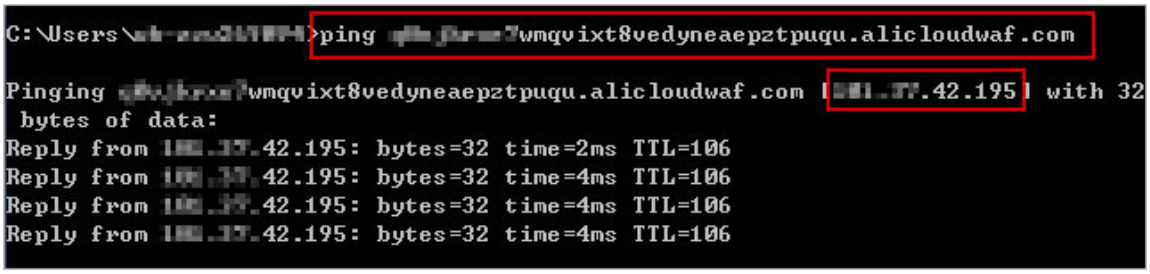
本地接入Anti-Bot

通过修改本地hosts文件（什么是hosts文件）模拟接入Anti-Bot，将从本地访问被防护站点的请求导向Anti-Bot实例。以Windows操作系统为例：

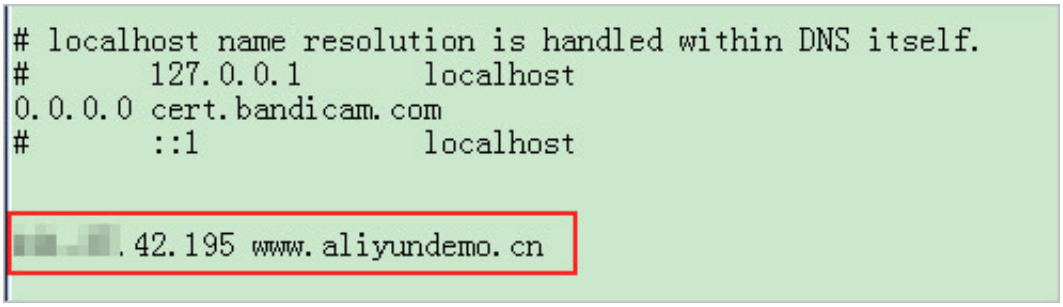
1. 用记事本或notepad++等文本编辑器打开hosts文件，hosts文件一般位于`C:\Windows\System32\drivers\etc\hosts`路径。
2. 在最后一行添加如下内容：`Anti-Bot实例的IP 被防护的域名`。

   以域名`www.aliyundemo.cn`为例，该域名已按照步骤1#添加域名添加到Anti-Bot实例中，且Anti-Bot为其分配了以下CNAME值：`xxxxxxxxxxxxxxx.alicloudwaf.com`

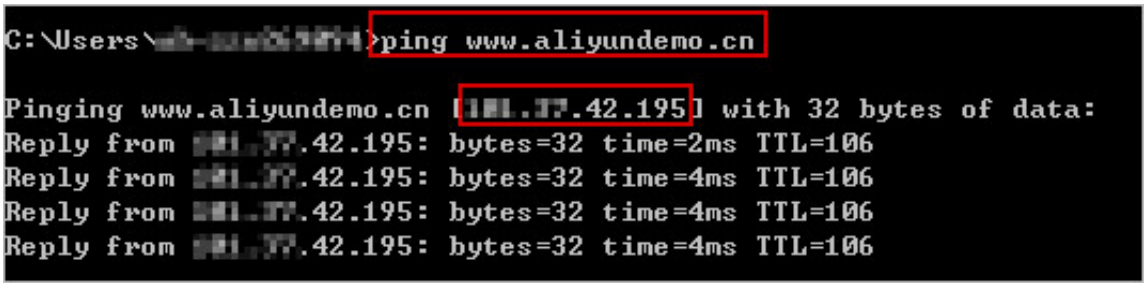
   a. 在Windows中打开cmd命令行工具，运行`ping xxxxxxxxxxxxxxx.alicloudwaf.com`命令获取Anti-Bot实例的IP。如下图所示，在响应结果中可以看到用来防护您的域名的Anti-Bot实例的IP。

   b. 在hosts文件添加如下内容，前面的IP地址即上一步获取的Anti-Bot实例的IP地址，后面的域名即被防护的域名。

3. 修改hosts文件后保存，然后本地对被防护的域名执行`ping`命令。



预期此时解析到的IP地址应该是刚才绑定的Anti-Bot实例的IP地址。如果依然是源站地址，可尝试刷新本地的DNS缓存（Windows的cmd下可以使用`ipconfig /flushdns`命令）。

验证Anti-Bot转发正常

确认hosts绑定已经生效（域名已经本地解析为Anti-Bot实例的IP）后，打开浏览器，输入该域名进行访问，如果Anti-Bot的域名接入配置正确，预期网站能够正常打开。

上述验证通过后，您可以参照*步骤4#修改DNS解析*，把业务流量正式切换至Anti-Bot上。

4 步骤4：修改DNS解析

通过修改域名的DNS解析记录将网站域名解析到爬虫风险管理（Anti-Bot Service，简称Anti-Bot）实例，完成业务正式接入。

建议您使用CNAME方式接入Anti-Bot，成功添加域名到防爬风险管理控制台后，Anti-Bot会为域名分配一个CNAME值，您只需添加/修改域名的CNAME解析记录为分配的CNAME值，即可完成接入。

注意:

参考步骤1#添加域名中的方法查看并记录Anti-Bot为该网站域名分配的CNAME值，如`xxxxxxxxxxxxx.alicloudwaf.com`。

CNAME接入说明

Anti-Bot通常采用CNAME解析的方式将网站接入进行防护，也支持A记录解析的方式接入。

注意:

强烈建议您采用CNAME解析的方式。因为在某些极端情况下（如节点故障、机房故障等），通过CNAME解析方式接入Anti-Bot，可以实现自动切换节点IP甚至将直接解析切回源站，从而最大程度保证业务的稳定运行，提供高可用性和灾备能力。

如果必须使用A记录解析的方式接入（例如，@记录与MX记录冲突等情况），您可以通过Ping Anti-Bot分配的CNAME值获取所分配的Anti-Bot实例的IP（该IP地址一般不会频繁变更），然后将网站域名通过A记录解析的方式接入Anti-Bot进行防护。

域名主机记录说明

关于域名的主机记录，以域名`abc.com`为例：

- www： 用于精确匹配www开头的域名，如`www.abc.com`，但无法匹配`abc.com`
- @： 用于匹配直接访问`abc.com`的情况
- *： 用于匹配泛域名，即匹配任意域名，如`blog.abc.com`、`www.abc.com`、`abc.com`等

修改DNS解析记录的注意事项

- 对于同一个主机记录，CNAME解析记录值只能填写一个，您可以将该记录值修改为Anti-Bot的CNAME值，从而将网站域名接入Anti-Bot进行防护。

- 由于不同DNS解析记录类型存在冲突，对于同一个主机记录，CNAME记录与A记录、MX记录、TXT记录等其他记录互相冲突。您需要删除原其它记录后再添加CNAME记录解析记录，或者将原记录类型修改为CNAME类型。

  关于DNS解析记录互斥的详细说明，参考解析记录冲突的规则。

  > **注意:**
  > 删除其它解析记录并新增CNAME解析记录的过程应尽可能在短时间内完成。如果删除A记录后长时间没有添加CNAME解析记录，可能导致域名无法正常解析。

- 如果必须保留MX记录（邮件服务器记录），您可以使用A记录解析的方式将域名解析到Anti-Bot的IP。通过Ping CNAME获取所分配的Anti-Bot实例的IP（该IP地址一般不会频繁变更）后，将网站域名A解析记录类型的记录值修改为该Anti-Bot实例的IP。

  > **注意:**
  > 使用A记录解析的方式进行接入，Anti-Bot将无法支持自动故障集群调度和故障bypass操作。

操作步骤

本文以阿里云云解析DNS和花生壳为例，介绍DNS配置方式，其它DNS提供商可以参考本文档进行类似配置。

阿里云云解析配置示例

> **注意:**
> 如果您使用阿里云云解析*DNS*进行域名解析，并且在执行步骤*1#*添加域名前已经为域名设置并启用了A记录（且如果是中国大陆域名，已完成备案），则您在添加网站配置时，可以一键添加网站并自动更新解析记录。以下操作步骤适用于您的域名已经添加到Anti-Bot域名列表，但DNS解析状态为异常的情况。

参照以下步骤，在阿里云云解析修改CNAME记录来接入Anti-Bot：

1. 登录阿里云云解析控制台。
2. 选择要操作的域名，单击其操作列的**解析设置**。

3. 选择要操作的主机记录，单击其操作列下的修改。

> **注意:**
>
> 您也可以删除已有的A记录，然后单击添加记录，添加一条新的CNAME记录。删除原解析记录后，请尽快添加CNAME解析记录，否则可能导致网站域名解析失败。

4. 将记录类型修改为**CNAME**，记录值修改为Anti-Bot所分配的CNAME值。

> **注意:**
>
> TTL值一般建议设置为600秒（即10分钟）。TTL值越大，则DNS记录的同步和更新越慢。

花生壳配置示例

部分域名提供商（如花生壳）可能不支持直接修改已有解析记录的记录类型和主机记录，您需要删除原有的A记录后，再添加新的CNAME解析记录。

注意:

删除原解析记录后，请尽快完成添加CNAME解析记录，否则可能导致网站域名解析失败。

参照下图设置来修改您的域名在花生壳上的DNS解析信息。

验证DNS配置

将网站域名的DNS解析切换至Anti-Bot后，您的网站域名即接入Anti-Bot进行防护。解析记录配置完成后，您可以通过Ping网站域名的方式或*17ce*等其他工具验证DNS解析生效情况。

注意:

由于DNS解析记录生效需要一定时间，如果本地测试域名无法访问，您可以等待10分钟后重新检查。