

<https://lz1y.github.io/2018/07/18/Punycode/>

## Internationalized Domain Name

由于互联网起源于美国,使得英文成为互联网上资源的主要描述性文字。这一方面促使互联网技术和应用的国际化,另一方面,随着互联网的发展特别在非英文国家和地区由此,[国际化域名](#)也应运而生。

国际化域名 ( 英语 : Internationalized Domain

Name , 缩写 : IDN ) 又称特殊字符域名 , 是指部分或完全使用特殊的文字或字母组成的互联网域名 , 包括法语、阿拉伯语、中文、斯拉夫语、泰米尔语、希伯来语或拉

虽说中文域名也已存在了十余年,但是它在中国互联网中却很少亮相。

一些公司,教育机构所使用的中文域名:

■■■■■: ■■■.■■■

■■■■■: ■■■■■.cn

除了中文域名外,还有诸如 .■■■ 这类中文顶级域名可供注册。

域名首页

英文域名

中文域名

.com .net .cn

.biz .cc .tv

.中国 .公司 .网络

新顶级域名

域名交易

域名预订

域名转入

域名增值服务

云解析

其他服务



全球最受欢迎的中文顶级域名!

99元/首年 [查看多年价格](#)

.com

立即查询

[查看域名注册规则](#)

### 关于.com域名

com为company简称,表示公司企业。.com是目前国际最广泛所有国际化公司都会注册.com域名。

▲ 2016年7月18日8点起, .com/.net域名注册成功后必须进行域名实名认证,否则域名将处于Serverhold状态,无法正常使用。[展开](#)

[详情](#)

全球注册量第一的域名

.com域名目前在全球的注册量已超过1.1亿,是全球最流行的域名。

最具声誉的域名

世界上众多知名公司都选择.com域名,利用.com进行网站宣传。

最广泛流行的域名

.com、.net、.org同为目前国际最广泛流行的通用域名格式。现全球用户超过1.1亿个。

极具资历的域名

.com是最早出现的域名后缀之一,极具资历,在网络上具有良好的信誉。

先知社区

而目前绝大部分的主流浏览器(Safari,chrome,Firefox等)也早已支持IDN。

```
lzy@47:~$ curl '贴吧。公司' -v
* Rebuilt URL to: 贴吧。公司/
* Input domain encoded as `UTF-8'
* Trying 137.74.127.233...
* Connected to 贴吧。公司 (137.74.127.233) port 80 (#0)
> GET / HTTP/1.1
> Host: xn--4qrp14k.xn--55qx5d
> User-Agent: curl/7.47.0
> Accept: */*
>
< HTTP/1.1 301 Moved Permanently
< Cache-Control: public, must-revalidate, proxy-revalidate, max-age=3600
< X-Powered-By: PHP/5.5.9-1ubuntu4.25
< Location: http://tieba.baidu.com
< Content-type: text/html
< Content-Length: 0
< Date: Wed, 18 Jul 2018 08:06:39 GMT
< Server: lighttpd/1.4.33
<
* Connection #0 to host 贴吧。公司 left intact
```



cURL的提示信息:

Input domain encoded as 'UTF-8'

cURL对域名做了如下转换:

■■■■■ => xn--4qrp14k.xn--55qx5d

想知道以上的转换是如何做到的,就不得不谈一下Punycode了.

## Punycode

Punycode (译: 域名代码) 是一种表示Unicode码和ASCII码的有限的字符集。例如: “münchen” (德国慕尼黑) 会被编码为“mnchen-3ya”。

Punycode的目的是在于国际化域名标签 (IDNA) 的框架中, 使这些 (多语言) 的域名可以编码为ASCII。编码语法在文档[RFC3492](https://tools.ietf.org/html/rfc3492)中规定。

Punycode is a simple and efficient transfer encoding syntax designed for use with Internationalized Domain Names in Applications (IDNA). It uniquely and reversibly transforms a Unicode string into an ASCII string. ASCII characters in the Unicode string are represented literally, and non-ASCII characters are represented by ASCII characters that are allowed in host name labels (letters, digits, and hyphens). This document defines a general algorithm called Bootstring that allows a string of basic code points to uniquely represent any string of code points drawn from a larger set. Punycode is an instance of Bootstring that uses particular parameter values specified by this document, appropriate for IDNA.

说白了,Punycode就是将Unicode字符串转成ASCII范围的字符,而xn--就是声明后面的字符串全部是Unicode编码。

## 安全相关

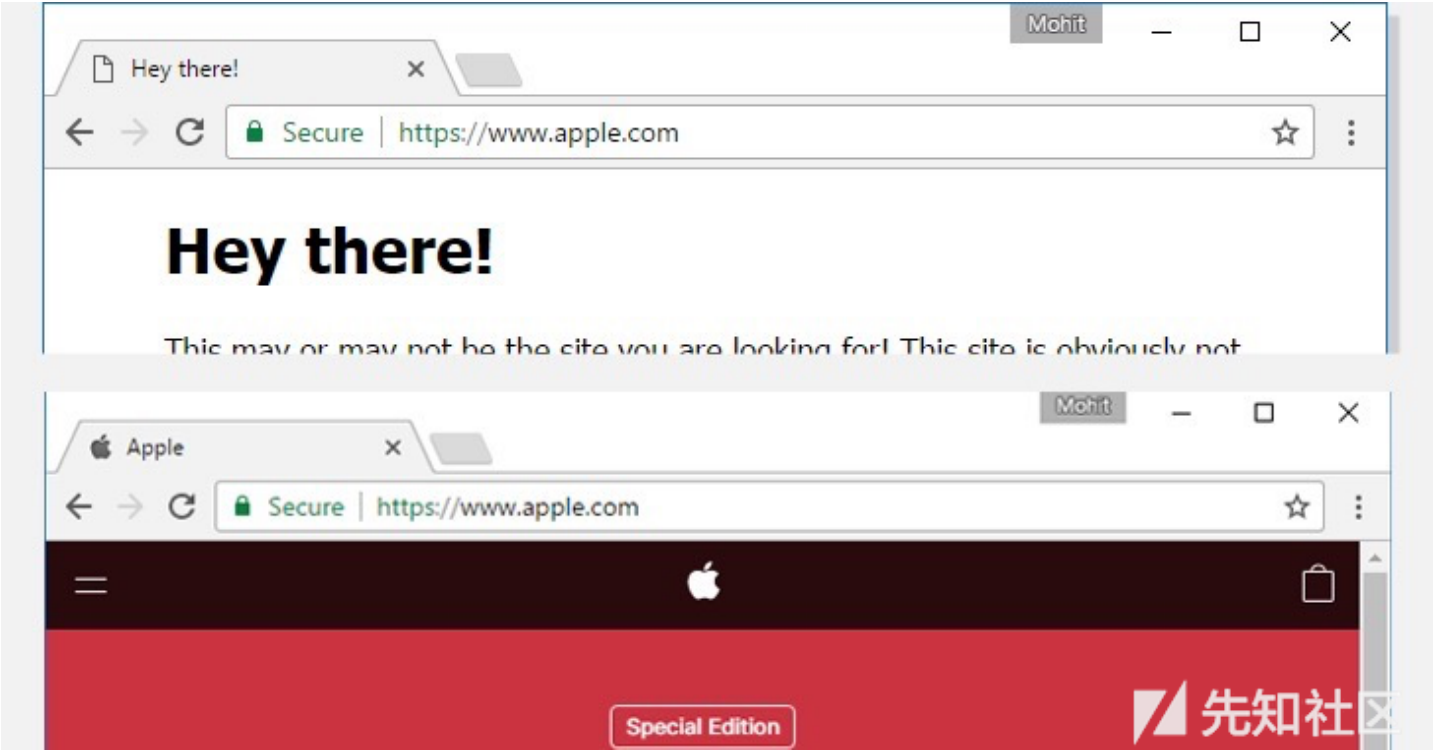
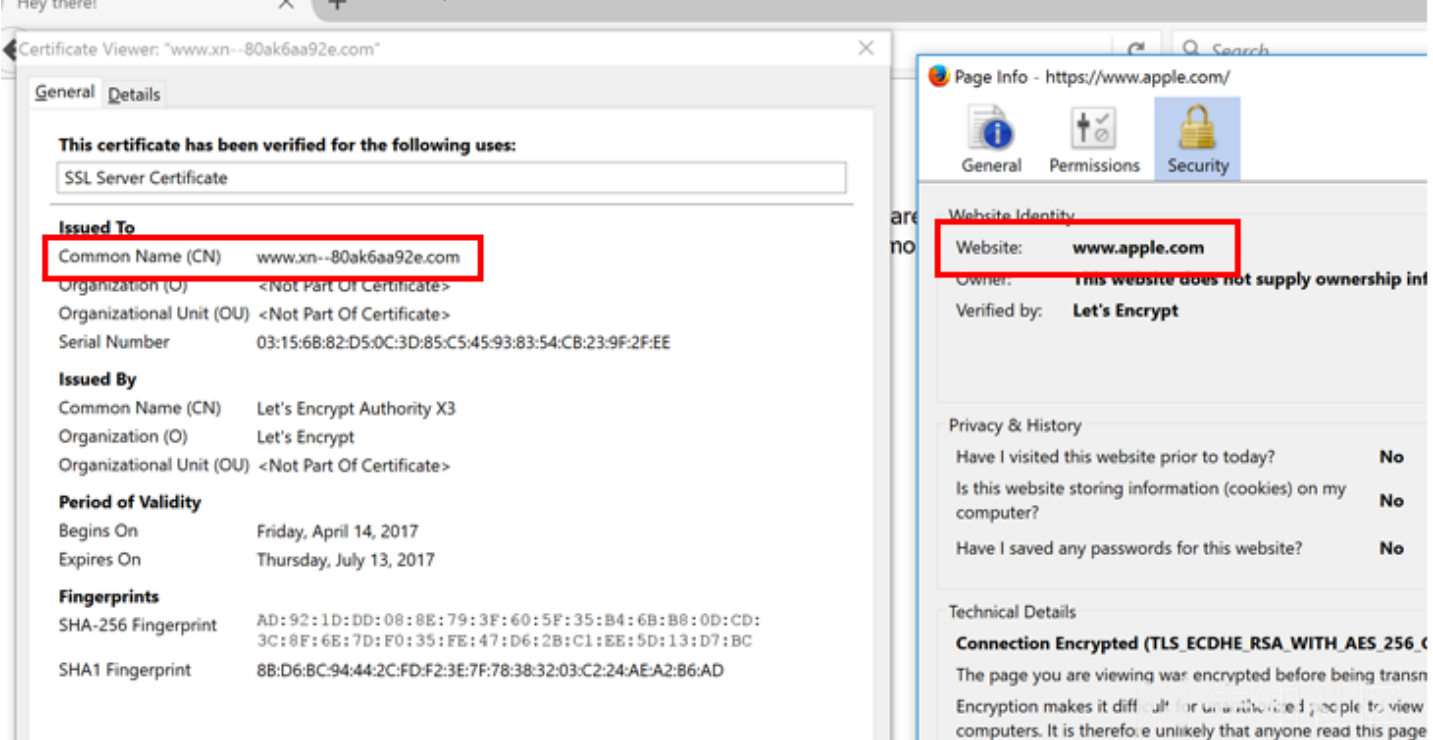
由于全世界语种繁多,各式各样的文字涌入了域名难免会发生一些问题.

### Punycode钓鱼攻击

许多Unicode字符, 代表的是国际化的域名中的希腊、斯拉夫、亚美尼亚字母, 看起来跟拉丁字母一样, 但是计算机却会把他们处理成完全不一样网的网址。

比如说, 斯拉夫字母“а” (U+0430) 和拉丁字母“a” (U+0041) 会被浏览器处理成不同的字符, 但是在地址栏当中都显示为“a”。

由于之前的部分浏览器,并不是在地址栏上显示Punycode编码后的域名,这样一来,就会产生很多令人混淆不清的域名.



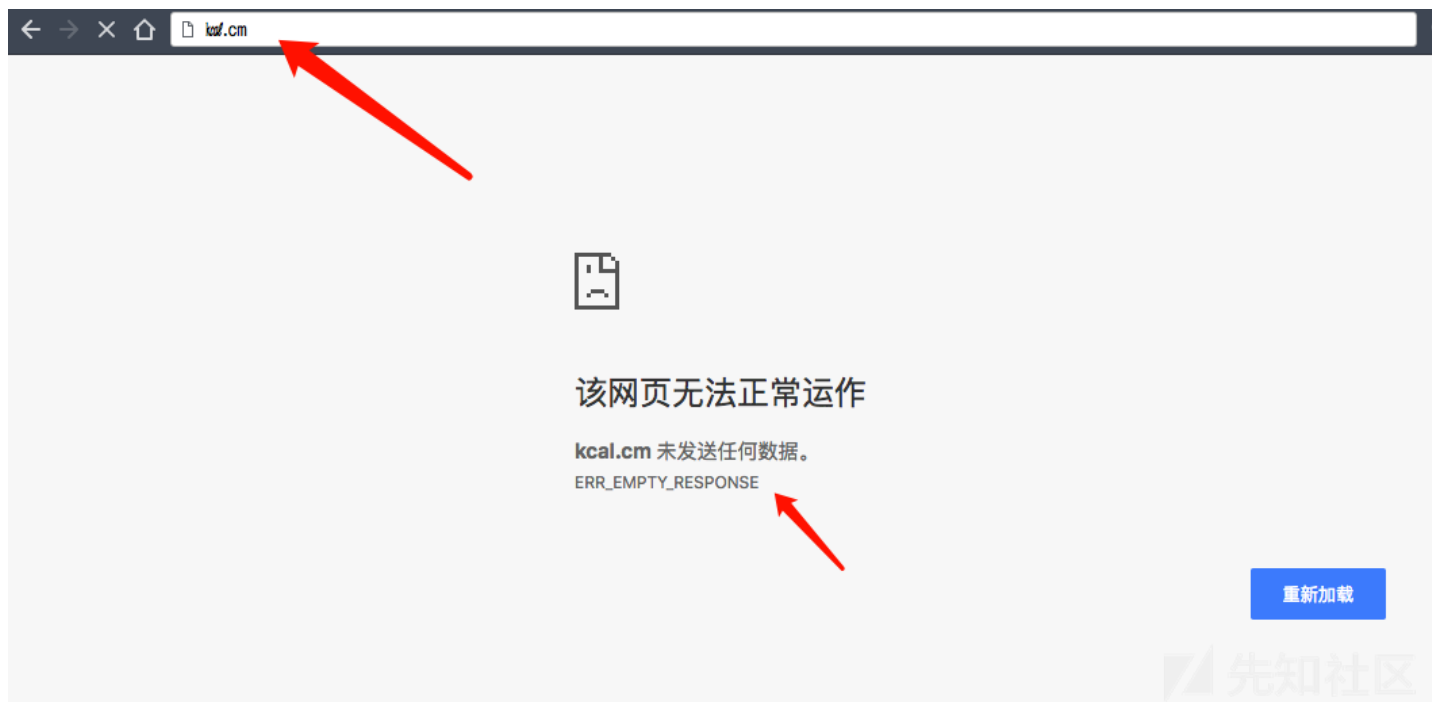
这个漏洞影响没有多长时间,在漏洞披露出来后没多久,受影响的厂商都在后来的更新把漏洞给修复了.

短域名

因为海量的Unicode字符,也经常出现一些比较有意思的事情.  
mramydneil师傅14年在乌云上就发过一篇文章[短域名进化史](#),详细讲解了利用Punycode低成本缩短域名至两位字符甚至一位字符.(ps:现在利用这个方法,长度至少三位)

例如:

```
■ ■ ■  
≥ 'kool.cm'.length  
◀ 3
```



#### 差异化解析

- JS中,编码前后的字符串不相同,但是同域

```
> document.domain
```

```
< "www.15.ee"
```

```
> 'www.15.ee' == document.domain
```

```
< false
```

```
> document.domain = '15.ee'
```

```
< "15.ee"
```

```
> document.domain = '15.eee'
```

```
✖ ▶ Uncaught DOMException: Failed to set the 'domain' property on 'Document': '15.eee' is not a suffix of '15.ee'. VM455:1
    at <anonymous>:1:17
```

- Bypass ssrf check

由于curl也支持IDN, 可以进行Punycode编码, 所以我们可以用来绕过日常的ssrf等漏洞的利用限制。

punycode解码: □ => 12

```
ziyi.liu@localhost ➜ nc -l -v 12121
```

```
GET / HTTP/1.1
```

```
Host: 0.0.0.1:12121
```

```
User-Agent: curl/7.54.0
```

```
Accept: */*
```

```
Punycode is amazing!
```


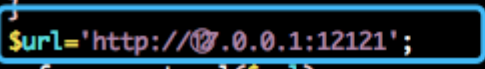
```
✖ ziyi.liu@localhost ➜ /tmp curl 0.0.0.1:12121
Punycode is amazing!
```

例如柠檬师傅曾经用过的一个[check ssrf脚本](#)。

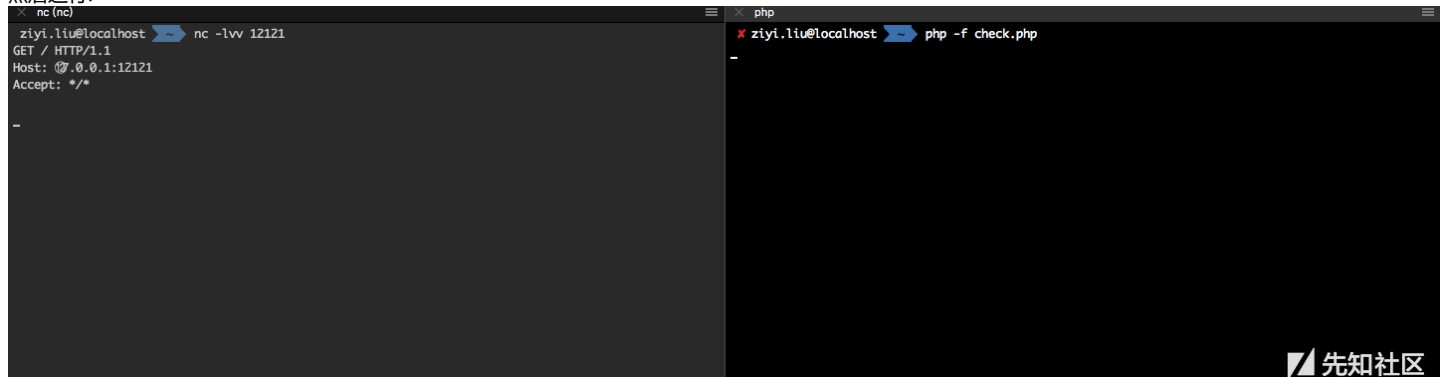
ps:柠檬师傅的文章[13th cuit\\_game\\_wp\\_web300\\_ssrf](#)

我们将其中的URL改成我们上面的URL:

```
34     {
35         $ch = curl_init();
36         curl_setopt($ch, CURLOPT_URL, $url);
37         curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
38         curl_setopt($ch, CURLOPT_HEADER, 0);
39         $output = curl_exec($ch);
40         $result_info = curl_getinfo($ch);
41         if ($result_info['redirect_url'])
42         {
43             safe_request_url($result_info['redirect_url']);
44         }
45         curl_close($ch);
46         print_r($output);
47     }
48
49 }
50 $url='http://0.0.0.1:12121';
51 safe_request_url($url);
52
```



然后运行:

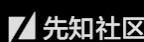


```
nc (nc)
ziyi.liu@localhost ➤ nc -lvv 12121
GET / HTTP/1.1
Host: 0.0.0.1:12121
Accept: */*

-

php
ziyi.liu@localhost ➤ php -f check.php
-

```

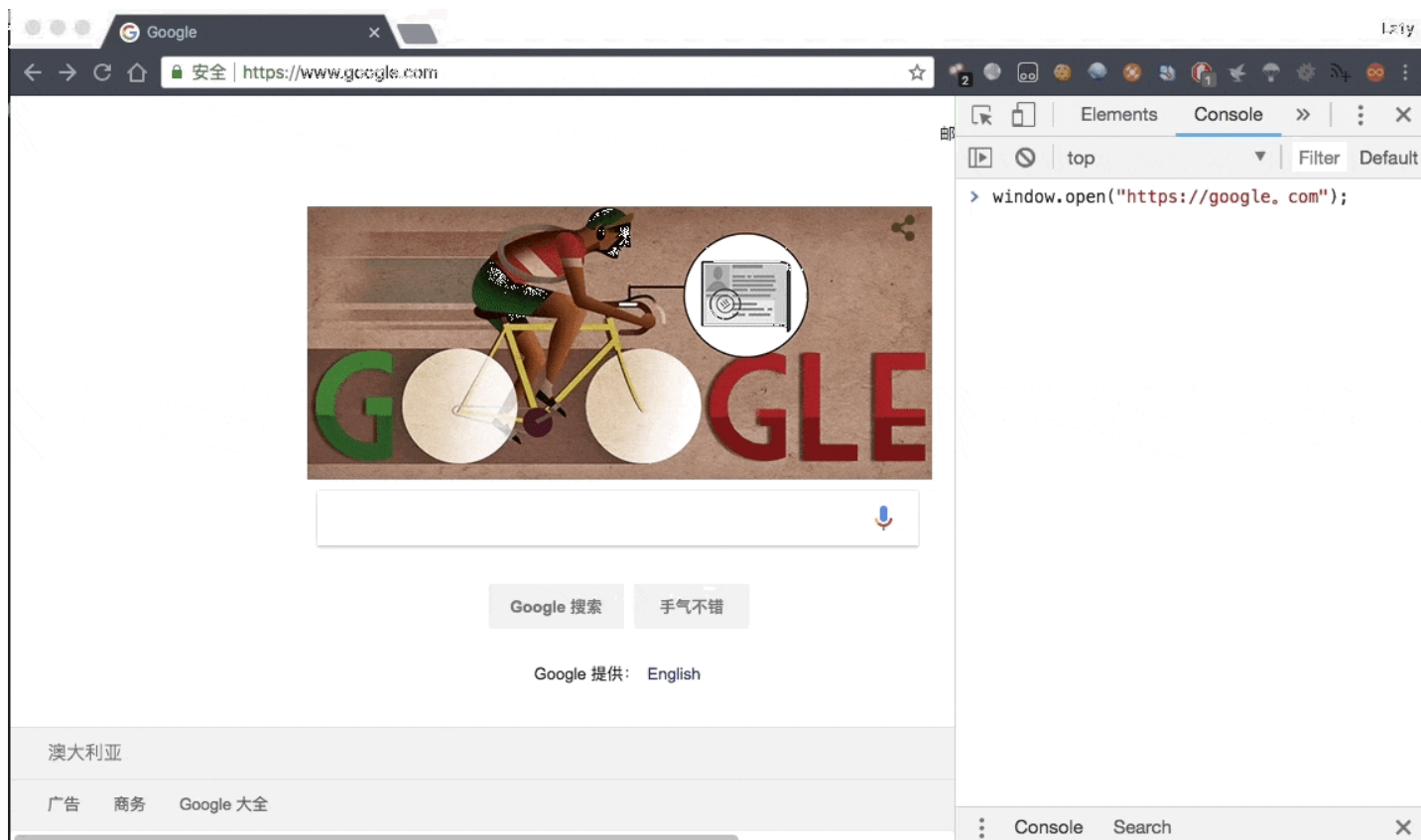


成功bypass了SSRF限制。

拓展

除了字符串,域名中的符号也是会参与到Punycode编码转换.

由于punycode不仅仅转换中文,所以除了简体中文的句号,还有台港澳的置中■,日本的半角中式句号■(U+FF61),台港澳置中■(U+FE52),中国大陆标准为靠左下■等...除了句号外,常见的符号还有破折号也有此类特性...



The end

至此，我只是非常浅显的分析，没有认真寻找漏洞案例。文章主旨也在于抛砖引玉...但是这个点的威力肯定不仅限于文中几点。希望有想法的读者可以跟我一起讨论讨论，扯Orz...

ref

- [rfc3492](#)
- [Punycode](#)
- [国际化域名](#)
- [短域名进化史](#)
- [xn-on-domain-what-it-means](#)
- [Internationalized domain name](#)
- [This Phishing Attack is Almost Impossible to Detect On Chrome, Firefox and Opera](#)

点击收藏 | 0 关注 | 1

[上一篇：JavaScript原型链污染](#) [下一篇：Ramnit代理服务器网络](#)

1. 0 条回复

- 动手手指，沙发就是你的了！

[登录](#) 后跟帖

先知社区

[现在登录](#)

热门节点

[技术文章](#)

[社区小黑板](#)

目录

[RSS](#) [关于社区](#) [友情链接](#) [社区小黑板](#)