

H3C SecPath IPS/ACG/UTM 中的URL地址过滤的配置说明

1. URL过滤说明



域名是

必选配置项，URI是可选配置项。它们又分别有固定字符串和正则表达式两种配置方式，因此总共有六种配置方法。

- 1) 域名——固定字符串方式：
- 2) 域名——正则表达式方式：
- 3) 域名——固定字符串方式 + URI——固定字符串方式：
- 4) 域名——固定字符串方式 + URI——正则表达式方式：
- 5) 域名——正则表达式方式 + URI——固定字符串方式：
- 6) 域名——正则表达式方式 + URI——正则表达式方式：

注意：用户所配置的“域名”将精确匹配HTTP请求中的Host字段后的完整内容，“URI”将精确匹配HTTP请求中的Get字段后的完整内容。如果选择固定字符串方式而且只截取其中一部分，将不会匹配成功，这一点和之前版本和主线版本是不同的。后续主线版本会同步修改过来。URI的内容从第一个左斜杠开始，注意不要遗漏。如果只想配置包含某个字符串的URI，须使用正则表达式进行通配。下面就六种配置方法分别示例如下：

- 1) 域名：www.sina.com

使用固定字符串方式，过滤域名为www.sina.com的所有HTTP请求

- 2) 域名：(news|sports)\.sina\.com\.cn

使用正则表达式方式，过滤域名news.sina.com.cn和sports.sina.com.cn

注意：不要忘记用右斜杠将点号转义，否则点号在正则表达式中的通配含义将引起误报。

- 3) 域名：www.sina.com URI：/index=1.html

均采用固定字符串方式，过滤如下HTTP请求：

http://www.sina.com/index=1.html

- 4) 域名：www.sina.com URI：.*badthing.*

域名采用固定字符串方式，URI采用正则表达式方式，过滤www.sina.com下所有包含badthing的资源网页

- 5) 域名：.*sina.* URI：/index.php

域名采用正则表达式，URI采用固定字符串方式，过滤所有主机名包含sina而且URI是/index.php的HTTP请求

- 6) 域名：.*sina.* URI：.*badthing.*

均采用正则表达式方式，过滤所有主机名含有sina而且URI含有badthing的HTTP请求

说明：

- ★目前的URL过滤是大小写不敏感的；
- ★目前的URL过滤是不支持中文关键词过滤的；
- ★附抓包信息中的URL过滤配置参考：

HOST：

(?is)Host: %s \r\n

URI：

(?i)^(GET|POST|PUT|HEAD|DELETE|TRACE|OPTIONS) [t]%s HTTP

%s表示用户在web中相应的输入(参考下图红色框中内容)



注意：URL地址过滤中规则匹配先后的顺序跟带宽管理中规则

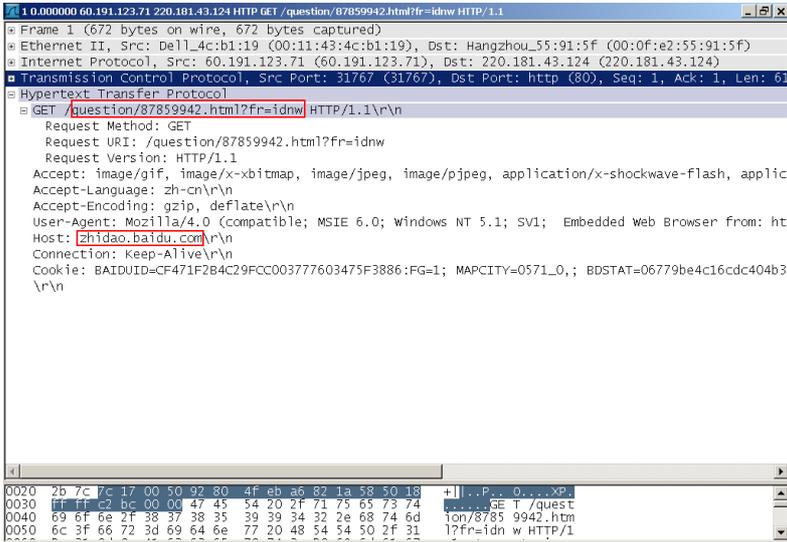
的顺序类似，即按照精度优先匹配：

域名&UR中正则表达式描述的网址的内涵越大（对应外延越小）的规则，也即描述的网址越精确，那么该条规则越优先匹配。换句话说默认规则肯定是最后被匹配的规则。

2. 应用举例

下面分情况，列出四种URL访问需求的实现。

1. 禁止所有web访问



ault前面加，否则达不到8个字符串，弹出下面的错误



2. 允许访问sina.com.cn域名下的所有资源

所谓sina.com.cn域名下的所有资源是指类似http://www.sina.com.cn/、http://news.sina.com.cn等其他各类主机资源。



域名中的

正则表达式为“*.sina.com.cn”，表示含有结尾部分为.sina.com.cn字符串的任意域名，这样二级域名如sports.sina.com.cn就可以正常访问。

3. 只允许访问www.sina.com.cn主机下的所有网页，不允许访问类似于sports.sina.com.cn等主机下的网站。

依据采用正则表达式还是固定字符串进行域名过滤，分为两种实现方式：域名过滤可以使用正则表达式“www.sina.com.cn”或用字符串方式表示为www.sina.com.cn，如下：



或者

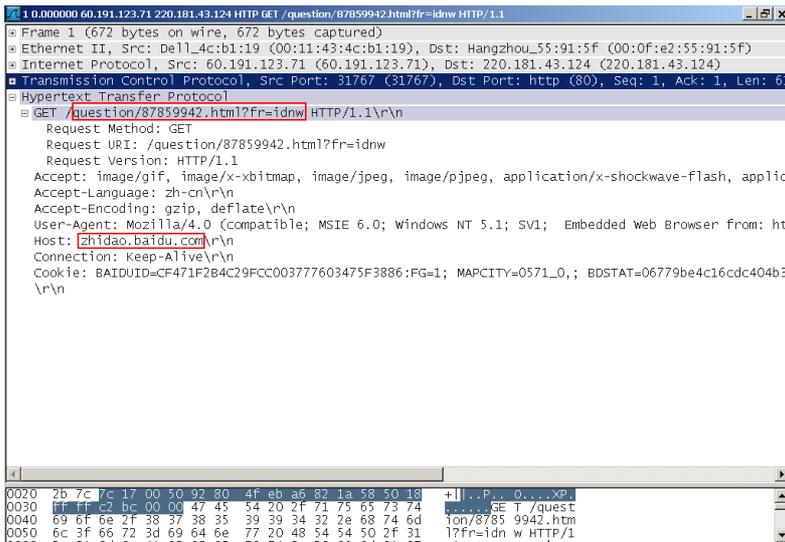


禁止访

问sports.sina.com.cn的配置和上面类似，不再列举。

4. 禁止访问URL地址中（不包含主机字段）包含world的任意网址

类似<http://x.x.x/world/> 的URL访问都要被过滤掉，可以通过如下方式实现：



因为域名

为任意字段，故可用正则表达式“.*.*.*”表示（规避最少8个字符的要求），URL路径含有world字段可以表示为“.*world.*”的正则表达式。

附录：

正则表达式介绍见附件