解决PPPoE宽带上网IPv6卡顿问题

Published on Aug 16, 2023 UTC by Wang Ye <https://wangye.org>

早期电信家用宽带支持IPv6的时候，我就尝试通过PPPoE拨号获取原生（Native）IPv6，一直使用正常，唯一遗憾就是由于当时IPv6刚刚起步，所以大部分支持IPv6的网站延迟都比较高，相应的访问速度比较慢，经过这几年发展国内IPv6业务日趋成熟，延迟和速度都有了较大提高，新的计算机、移动设备等都能较好的支持IPv4和IPv6双栈（Dual-Stack）访问。

近两年开始出现部分网站访问卡顿、应用无法加载的情况，经过调查发现这些网站和应用使用了IPv6的访问方式，极大影响了浏览体验，于是我决定先暂停IPv6的解析（DNS仅返回A记录）以缓解卡顿的问题，不过后来搜索网络我终于找到了问题的根源，彻底解决了，在这里把解决的过程记录下来以备忘用，如果读者想快速解决问题的话建议直接跳转到第2节 配置路由器TCP MSS。

1 配置返回IPv4 Only DNS服务器

可能有人说配置返回IPv4 Only的DNS服务器不如直接关闭路由器的IPv6拨号功能，当然我这边只是缓解，在尽量不影响路由器IPv4和IPv6双拨前提下的，也许你可能需要IPv6，比如某些业务场景、外网穿透等等，又或者你无法控制拨号路由器。

返回IPv4 Only也就是说不解析AAAA记录，目前我内网所部署的DNS服务器使用了dnsmasq提供解析服务，我搜索了网络和相关文档，dnsmasq无配置“仅返回IPv4的A记录”或者“过滤IPv6的AAAA记录”的选项，如果想继续保留dnsmasq提供解析服务且完成这类配置，那么临时的解决方案就是增加支持该配置的上游DNS服务器。

1.1 使用bind9作为上游DNS服务器

期间考察了dnscrypt-proxy、unbound和老牌的bind9，其中dnscrypt-proxy和bind9是支持这类功能，dnscrypt-proxy原生支持，bind9需要安装插件，本节介绍如何配置bind9支持，以及将bind9作为dnsmasq的上游DNS服务器程序。

以Alpine操作系统为例，首先安装bind9和plugins：

```
apk add --no-cache --update bind bind-plugins
```

编辑named.conf.options配置文件，在配置文件最开始添加如下配置节：

```
plugin query "/usr/lib/bind/filter-aaaa.so" {
  filter-aaaa-on-v4 yes;
  filter-aaaa-on-v6 yes;
};
```

重启bind9服务，完成设置。

1.2 修改设备DNS服务器

现在我们将重新配置的dnsmasq服务器上线并让局域网终端使用此服务器进行DNS查询，需要注意的是我们必须限定终端只能通过我们自己配置的dnsmasq服务器进行DNS查询，否则仍然有可能返回IPv6地址导致缓解失效。

1.2.1 直接配置

这个不需要多介绍，一般联网设备的网络配置里面都可以手动指定DNS服务器地址，需要注意的是我们这里仅需要指派首要DNS服务器（Primary DNS Server）地址为我们dnsmasq的IP地址即可，如果指派了次要DNS服务器（Secondary DNS Server），那么DNS查询很有可能在这两个服务器之间随机发起，此时次要DNS如果没有进行相关配置那么仍然有可能得到IPv6地址。

1.2.2 路由转发

如果每个设备都配置DNS服务器那么就很不方便了，这里可以在路由器上做相关设置，比如我的这篇文章《iptables劫持并拦截DNS查询53端口实现转向(Redirect)》，又或者将DNS服务器设置为旁路设备，侦测过往数据包并进行拦截修改。

2 配置路由器TCP MSS

配置这么多直到看到这篇帖子《开启 IPv6 后网速变得很慢？可能是 PMTU 黑洞的问题》才了解到问题的根源，主要是本机与远端服务器路径存在 PMTU 黑洞问题导致包的大小一旦超过了 PMTU，会被无声地丢弃，直到 TCP 协议超时进行重传，从而带来传输缓慢的问题，楼主也提到其实这个问题IPv4也存在，只不过IPv4下路由器的 TCP 开启了 MSS Clamping，通过嗅探 TCP 握手包，把 MSS 值降低，从而避免了这个问题。

帖子楼主推荐对于通过 PPPoE 虚拟拨号建立宽带连接的，默认 MTU 是 1500，由于 PPPoE 隧道有 8 个 bytes 的开销，所以 PPPoE 虚拟连接的 MTU 就是 1500-8=1492，减掉 IPv4 包头（ 20 字节）和 TCP 包头（ 20 字节），可以得知 IPv4 下需要把 MSS 设为 1452 以下。而IPv6 的包头是 40 字节，所以 IPv6 下需要把 MSS 设为 1432 以下。

2.1 路由器TCP MSS配置命令

(1) iptables自动MSS，假设PPPOE虚接口是`pppoe0`：

```
# IPv4
iptables -t mangle -A POSTROUTING -p tcp --tcp-flags SYN,RST SYN -o
pppoe0 -j TCPMSS --clamp-mss-to-pmtu

# IPv6
ip6tables -t mangle -A POSTROUTING -p tcp --tcp-flags SYN,RST SYN -o
pppoe0 -j TCPMSS --clamp-mss-to-pmtu
```

(2) 手动指定MSS，假设PPPOE虚接口是`pppoe0`：

```
# IPv4
iptables -t mangle -A POSTROUTING -p tcp --tcp-flags SYN,RST SYN -o
pppoe0 -j TCPMSS --set-mss 1452

# IPv6
ip6tables -t mangle -A POSTROUTING -p tcp --tcp-flags SYN,RST SYN -o
pppoe0 -j TCPMSS --set-mss 1432
```

(3) firewall-cmd自动MSS：

```
# IPv4
firewall-cmd --permanent --direct --add-passthrough ipv4 -t mangle -I
FORWARD -p tcp --syn -j TCPMSS --clamp-mss-to-pmtu

# IPv6
firewall-cmd --permanent --direct --add-passthrough ipv6 -t mangle -I
FORWARD -p tcp --syn -j TCPMSS --clamp-mss-to-pmtu
```

(4) RouterOS里设置IPv6 MSS的命令：

其中pppoe-out1是WAN口，1420是要MSS值，请根据需要修改。

```
/ipv6 firewall mangle add chain=forward out-interface=pppoe-out1
protocol=tcp tcp-flags=syn action=change-mss new-mss=1420
```

RouterOS支持自动MSS，所以推荐直接使用下面的命令（WAN口仍然为pppoe-out1）。

```
/ipv6 firewall mangle add action=change-mss chain=forward new-
mss=clamp-to-pmtu out-interface=pppoe-out1 passthrough=yes
protocol=tcp tcp-flags=syn
```

(5) UBNT Edgerouter 系列的IPv6 MSS设置方法：

```
set firewall options mss-clamp6 interface-type pppoe
set firewall options mss-clamp6 mss 1420
```

3 总结

随着IPv6的全面铺开，越来越多的网站或者网络应用支持IPv6，同时启用IPv4和IPv6双栈可以一定程度缓解网络拥塞，但如果IPv6配置不当反而会拖累整个网络。

External References (Links)

《iptables劫持并拦截DNS查询53端口实现转向(Redirect)》
https://wangye.org/blog/archives/956/
《开启 IPv6 后网速变得很慢？可能是 PMTU 黑洞的问题》
https://www.v2ex.com/t/800024
firewall-cmd自动MSS
https://firewalld.org/2020/10/tcp-mss-clamp