


How to run multiple Tor/Privoxy processes for Scrapy with different ips at the same time?

Asked 4 years, 4 months ago

Modified 2 years, 3 months ago

Viewed 3k times



Can someone please explain me how to make it work? I want to run multiple tors at the same time via privoxy for my scrapy project... Each one should be with a different ip adress.

0



proxy

scrapy

tor

privoxy



Share

Improve this question

Follow



asked Nov 14, 2020 at 15:48



 EXOR6

11

1

1


Is privoxy a requirement? You can use Tor's [HTTPTunnelPort](#) directive to specify multiple HTTP proxy ports, each with stream isolation (different IP) and then configure scrapy's HTTP proxy middleware to use Tor's HTTP proxy. Each scrapy instance can use their own HTTPTunnelPort and they'll each have their own IP, all from a single Tor instance. – [drew010](#) Nov 15, 2020 at 1:37



1 Answer

Sorted by:

Highest score (default)





First of all sorry for my bad english cause it's not my native language (i'm Russian). I came across many "problems" while trying to set up a multiple instances of tor and privoxy so that each one can work with different ip for my scrapy. I'll show you how it's done.



I'll be working on Kali Linux 2.0 I'll be also working as root.



Step 1. Get tor and privoxy:



```
apt-get install tor
apt-get install privoxy
```



Step 2. Make tor duplicates...

```
cp /var/lib/tor -r /var/lib/tor2
cp /var/lib/tor -r /var/lib/tor3
cp /var/lib/tor -r /var/lib/tor4
```



Step 3. Make torrc duplicates...

```
cp /etc/tor/torrc torrc2
cp /etc/tor/torrc torrc3
cp /etc/tor/torrc torrc4
```

Step 4. Now we gonna edit each torrc file, you can delete everything inside and input these configs:

```
SocksPort 9050 (9060, 9070, 9080)
ControlPort 9051 (9061, 9071, 9081)
DataDirectory /var/lib/tor (tor2, tor3, tor4)
```



Step 5. Let's test that everything works well, let's open 4 terminals and launch tor in everyone like this:

```
Terminal 1: tor -f /etc/tor/torrc
Terminal 2: tor -f /etc/tor/torrc2
Terminal 3: tor -f /etc/tor/torrc3
Terminal 4: tor -f /etc/tor/torrc4
```

Now open a new terminal and let's curl this site "[http://ipinfo.io/ip](#)", it's gonna give us our ip. Commands for each tor will be like this:

```
curl --proxy socks5h://localhost:9050 http://ipinfo.io/ip
curl --proxy socks5h://localhost:9060 http://ipinfo.io/ip
curl --proxy socks5h://localhost:9070 http://ipinfo.io/ip
curl --proxy socks5h://localhost:9080 http://ipinfo.io/ip
```

If you did everything well, each one should return a different ip. Now we have 4 instances of tor running in the same time. But tor uses sock5 proxy and for our scrapy project we need it to be http proxy. So we gonna plug privoxy.



Step 1. Let's copy our privoxy folder 3 times like with tor...

```
cp -a /etc/privoxy /etc/privoxy2
cp -a /etc/privoxy /etc/privoxy3
cp -a /etc/privoxy /etc/privoxy4
```

Step 2. Now we gonna edit each config file in each privoxy folder:

```
mousepad /etc/privoxy(2,3,4)/config
```

First we have to uncheck the "forward-socks5t" and change the port for each config file:

```
forward-socks5t  /          127.0.0.1:9050 .
forward-socks5t  /          127.0.0.1:9060 .
forward-socks5t  /          127.0.0.1:9070 .
forward-socks5t  /          127.0.0.1:9080 .
```

Also change the listen adress:

```
listen-address 127.0.0.1:8118
listen-address 127.0.0.1:8128
listen-address 127.0.0.1:8138
listen-address 127.0.0.1:8148
```

Step 3. We gonna make a copy of another privoxy folder:

```
cp /etc/init.d/privoxy /etc/init.d/privoxy2
cp /etc/init.d/privoxy /etc/init.d/privoxy3
cp /etc/init.d/privoxy /etc/init.d/privoxy4
```

Step 4. Edit the privoxy file in each folder:

```
mousepad /etc/init.d/privoxy(2,3,4)
NAME=privoxy(2,3,4)
OWNER=privoxy(2,3,4)
LOGDIR=/var/log/privoxy(2,3,4)
CONFIGFILE=/etc/privoxy(2,3,4)/config
```

Step 5. Finally, make copies for logs for each privoxy:

```
cp -a /usr/sbin/privoxy /usr/sbin/privoxy2
cp -a /usr/sbin/privoxy /usr/sbin/privoxy3
cp -a /usr/sbin/privoxy /usr/sbin/privoxy4
```

Done. Let's test, relaunch 4 tors in different terminals like in the step 5. Now we gonna launch in a different terminal 4 privoxys like this:

```
start-stop-daemon --start --exec /usr/sbin/privoxy --pidfile /var/run/privoxy.pid -- --user root /etc/privoxy/config
start-stop-daemon --start --exec /usr/sbin/privoxy2 --pidfile /var/run/privoxy2.pid -- --user root /etc/privoxy2/config
start-stop-daemon --start --exec /usr/sbin/privoxy3 --pidfile /var/run/privoxy3.pid -- --user root /etc/privoxy3/config
start-stop-daemon --start --exec /usr/sbin/privoxy4 --pidfile /var/run/privoxy4.pid -- --user root /etc/privoxy4/config
```

To test if it works let's curl the same site:

```
curl --proxy http://127.0.0.1:8118 http://ipinfo.io/ip
curl --proxy http://127.0.0.1:8128 http://ipinfo.io/ip
curl --proxy http://127.0.0.1:8138 http://ipinfo.io/ip
curl --proxy http://127.0.0.1:8148 http://ipinfo.io/ip
```

If it gives you different ips then you did everything well. If you wanna close all privoxys just type :

```
kill privoxy(2,3,4)
```

Now, when you gonna make spiders in scrapy, you can use a different proxy for each spider, here's an example:

```
import requests
from stem import Signal
from stem.control import Controller

def get_new_ip(2,3,4)():
    with Controller.from_port(port=9051(9061,9071,9081)) as controller:
        controller.authenticate()
        controller.signal(Signal.NEWNYM)
    response = requests.get('https://api.myip.com/', proxies={'https': '127.0.0.1:8118(8128,8138,8148)'})
```

```
print(response.text)
```

```
class ProxyMiddleware(2,3,4)(object):
    _requests_count = 0

    def process_request(self, request, spider):
        self._requests_count += 1
        if self._requests_count > 5:
            self._requests_count = 0
            get_new_ip1()

        request.meta['proxy'] = 'http://127.0.0.1:8118(8128,8138,8148) '
        spider.log('Proxy : %s' % request.meta['proxy'])
```

That's it, I hope that this gonna help you out!)) I used to struggle a lot to get all this work together and didn't find any tutorial which explains step by step how to make it work.

Good luck!)

-Mikhail Aleksandrovskiy (offshore47)

Share Improve this answer Follow

answered Nov 14, 2020 at 15:50

 **Mikhail Aleksandrovskiy**

51 3



Start asking to get answers

Find the answer to your question by asking.

Ask question

Explore related questions

- proxy
- scrapy
- tor
- privoxy

See similar questions with these tags.

