在搭建代理的时候, 由于我们的代理是为了隐藏爬虫, 所以必须使用http正向代理中的高匿代理, 否则会被网易云的后台发现异常. 这里使用的是自己使用Node.js编写的一个简单的匿名代理.

代理的原理是客户端先将请求发送到代理服务器, 由代理服务器向目标服务器发起请求, 并将请求结果发回客户端.

代码如下:

const PROXY\_PORT = 65535

const http = require('http')

const net = require('net')

const url = require('url')

function request(cReq, cRes)

{

const u = url.parse(cReq.url)

console.log(`[http.request] ${cReq.method} ${u.hostname}:${u.port || 80} ${u.path}`)

const options =

{

hostname : u.hostname,

port : u.port || 80,

path : u.path,

method : cReq.method,

headers : cReq.headers,

}

const pReq = http.request(options, function(pRes)

{

cRes.writeHead(pRes.statusCode, pRes.headers)

pRes.pipe(cRes);

}).on('error', function(e)

{

console.error(e)

cRes.end()

});

cReq.pipe(pReq)

}

function connect(cReq, cSock)

{

console.log(`[http.connect] ${cReq.url}`)

const u = url.parse('http://' + cReq.url);

const pSock = net.connect(u.port, u.hostname, function()

{

cSock.write('HTTP/1.1 200 Connection Established\r\n\r\n');

pSock.pipe(cSock);

}).on('error', function(e)

{

console.error(e)

cSock.end();

});

cSock.pipe(pSock);

}

http.createServer()

.on('request', request)

.on('connect', connect)

.on('error', function(err)

{

console.error(err)

})

.listen(PROXY\_PORT, function()

{

console.log(`proxy server online: http://localhost:${PROXY\_PORT}`)

});