Computer Networks Lab 7

CS F303

Susmit Wani 2018A7PS0116G

Lab 7 : TCP File Downloader

Dasgaonkar Yogesh Namdeo • 9:30 AM (Edited 11:41 AM)

Labs • 16 points **Due Apr 5, 6:59 PM**

Write TCP client to download and locally save a file.

1. The client, as a command-line argument, accepts the URL of the file. The URL can be HTTP or HTTPS-based. (2 marks)
2. If the URL does not exist or unreachable, the client prints an error and exits. (2 marks)
3. After connecting to the server, the client sends a GET request to the server. (2 marks)
4. The client downloads the whole file and saves it to the local folder with the same file name as mentioned in the URL. (8 marks)
5. If a file with the same name exists, the old file is overwritten. (1 mark)
6. The client exists after downloading the file. (1 mark)

Notes:
1. All the students will upload C code along with a PDF file containing the screenshots of the executed program on Google classroom
2. Submit a README file giving instruction for compilation and execution
3. You will get zero marks if the PDF or the README is missing
4. Please submit a single zip file named <Name>_<ID_number>. Make sure to turn in your submissions on time.

Compiling the program

```
susmit@susmit-VB:~/Desktop/CN Lab 7$ gcc -o client client.c -lssl -lcrypto
client.c: In function ‘main’:
client.c:140:24: warning: passing argument 1 of ‘strtok’ discards ‘const’ qualifier from pointer target type [-Wdiscarded-qualifiers]
        token = strtok(argv[1], s);
                       ^~~~
In file included from client.c:3:0:
/usr/include/string.h:335:14: note: expected ‘char * restrict’ but argument is of type ‘const char *’
 extern char *strtok (char *__restrict __s, const char *__restrict __delim)
              ^~~~~~
client.c:245:24: warning: passing argument 1 of ‘strtok’ discards ‘const’ qualifier from pointer target type [-Wdiscarded-qualifiers]
        token = strtok(argv[1], s);
                       ^~~~
In file included from client.c:3:0:
/usr/include/string.h:335:14: note: expected ‘char * restrict’ but argument is of type ‘const char *’
 extern char *strtok (char *__restrict __s, const char *__restrict __delim)
              ^~~~~~
```

As we can clearly see, we need to include the -lssl and the -lcrypto flags with our line. There are some warnings but we can ignore them for now. Use the -w flag to suppress warnings

Entering wrong website

```
susmit@susmit-VB:~/Desktop/CN Lab 7$ ./client http://www..ggwp.com/ggez.txt

URL: http://www..ggwp.com/ggez.txt
Protocol: http
Domain: www..ggwp.com
Path: ggez.txt
Website unavailable/unreachable.
ERROR CODE: -2susmit@susmit-VB:~/Desktop/CN Lab 7$
susmit@susmit-VB:~/Desktop/CN Lab 7$
```

We can see that the protocol, domain and the filename are identified. But website is unreachable so error message is shown and client exits. This satisfies one of the conditions in the question.

Running the Code

To run the code we add the website with http or https prefix as a command line input. Also, the code is made to download files so a file with the proper extension is needed at the end. Like we have .html here.

```
susmit@susmit-VB:~/Desktop/CN Lab 7$ ./client http://info.cern.ch/hypertext/WWW/TheProject.html

URL: http://info.cern.ch/hypertext/WWW/TheProject.html
Protocol: http
Domain: info.cern.ch
Path: hypertext/WWW/TheProject.html
Host Name->webafs706.cern.ch
IP ADDRESS->119.101.98.97
Socket created successfully.
connected to the server successfully
The send req is:GET /hypertext/WWW/TheProject.html HTTP/1.1
Host: info.cern.ch
Connection: close

Sending the GET request to the website
TheProject.html
size of p: 1127
File download successful!
File download successful!
File download successful!
susmit@susmit-VB:~/Desktop/CN Lab 7$
```

Running the Code (contd...)

As we can see in the previous slide, the domain name, and the path of the file are identified. The IP and the host name of the website are also identified.

After that, we can see that the socket is created and the connection to server is made. The GET request can be seen in the terminal window.

The terminal also gives us an indication if it was successful in retrieving the file.

The working is same for any https site as well. Screenshot attached in the next slide.

Before downloading I will make a blank file by the name of BIO_set_conn_Hostname.html to see if it overwrites the data or not. It actually does overwrite the file in the same folder and exits after executing the command.

So, all requirements in the pdf are fulfilled.

```
susmit@susmit-VB:~/Desktop/CN Lab 7$ > BIO_set_conn_hostname.html
susmit@susmit-VB:~/Desktop/CN Lab 7$ cat BIO_set_conn_hostname.html
susmit@susmit-VB:~/Desktop/CN Lab 7$
```

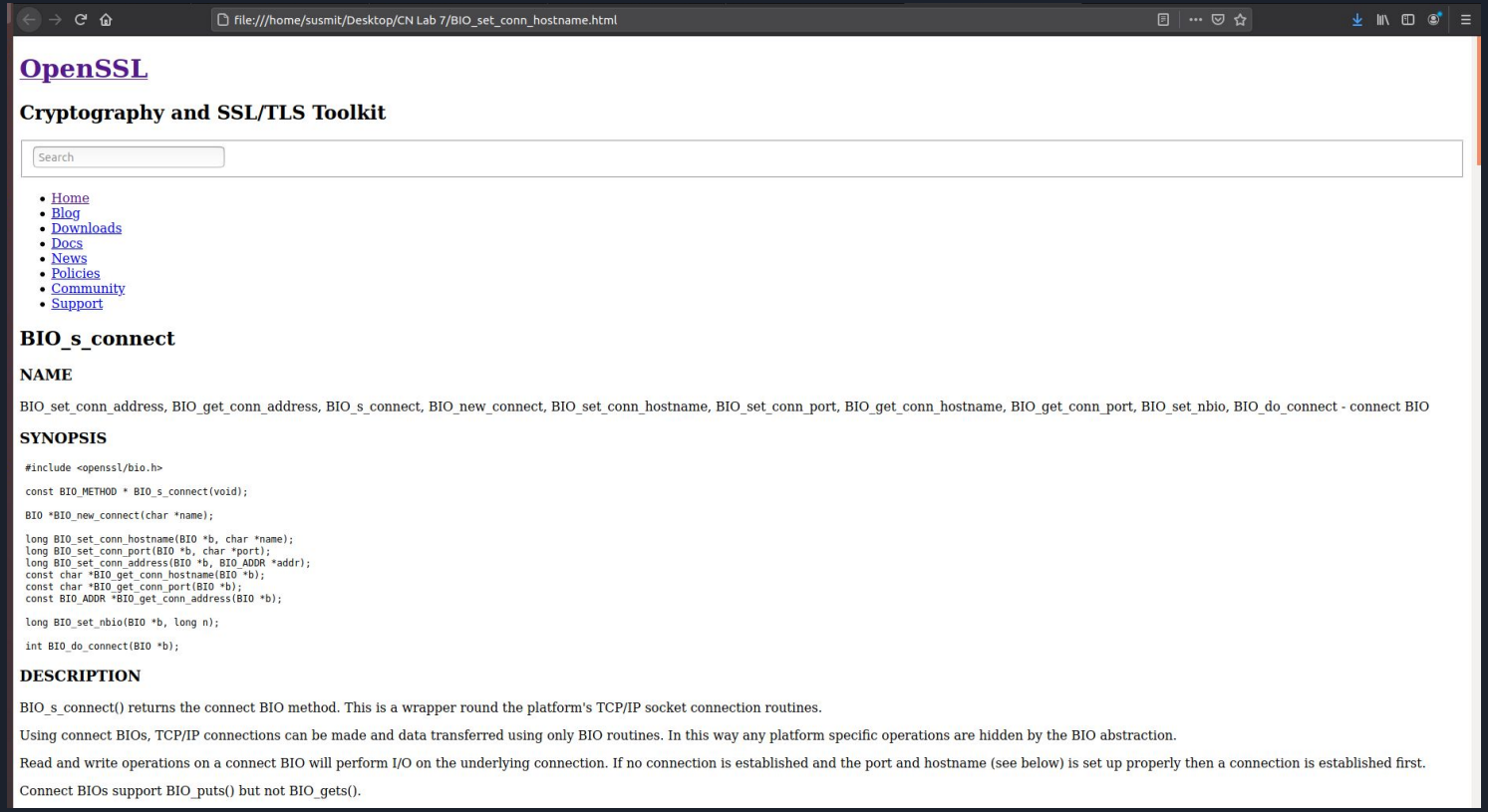
```
susmit@susmit-VB:~/Desktop/CN Lab 7$ > BIO_set_conn_hostname.html
susmit@susmit-VB:~/Desktop/CN Lab 7$ cat BIO_set_conn_hostname.html
susmit@susmit-VB:~/Desktop/CN Lab 7$ ./client https://www.openssl.org/docs/man1.1.0/man3/BIO_set_conn_hostname.html

URL: https://www.openssl.org/docs/man1.1.0/man3/BIO_set_conn_hostname.html
Protocol: https
Domain: www.openssl.org
Path: docs/man1.1.0/man3/BIO_set_conn_hostname.html
Host Name->e3102.dscx.akamaiedge.net
IP ADDRESS->101.51.49.48
hostname: www.openssl.org
Hostname set

URL: https://www.openssl.org/docs/man1.1.0/man3/BIO_set_conn_hostname.html
Protocol: https
Domain: www.openssl.org
Path: docs/man1.1.0/man3/BIO_set_conn_hostname.html
The send req is:GET /docs/man1.1.0/man3/BIO_set_conn_hostname.html HTTP/1.0
Host: www.openssl.org
Connection: close

Sending the GET request to the website
BIO_set_conn_hostname.html
size of p: 9687
File download successful!
File download successful!
susmit@susmit-VB:~/Desktop/CN Lab 7$ cat BIO_set_conn_hostname.html
<!DOCTYPE html>
<html lang="en">
<!-- OSSL: original subdir: crypto -->
<!-- OSSL: subdir: man3 -->
<!-- OSSL: section: 3 -->
<!-- OSSL: description: connect BIO  -->
```

Screenshot of https code output

As we can see, initially the file is empty but later it is overwritten by the new file.

The html file is shown alongside

Note that it is download using https



Limitations

The html, txt and pdf files can be downloaded but there are some issues in downloading images. Essentially in the formatting, there is some problem. The files are downloaded but we can't view them due to some formatting issues. Upon checking the size, we can see that entire file is downloaded. So I believe there is some error in my code while taking care of the header message received along with file contents.

That being said, the code is still functional and can be tested on files like json, html and txt and pdfs.