

FTP based File Transfer Application using Sockets

Sbonelo Mdluli : 1101772

School of Electrical & Information Engineering, University of the Witwatersrand, Private Bag 3, 2050, Johannesburg, South Africa

Abstract—This report presents an implementation of a file transfer application using TCP. The development of the application is per RFC959 which presents the standard for file transfer application implementation. The implementation contains multiple commands and a GUI to be used by the client. A multi threaded version of the application is also implemented to allow multiple user to connect to the server. The implementation can be further improved to support bulk transmission and to allow the user to have GUI version of the server. Passive mode can also be improved to support more multimedia file formats. Wireshark is used to verify the functionality of the implemented application.

I. INTRODUCTION

File transfer protocol (FTP) is a computer network protocol used for transferring files between computers . FTP uses Transfer Control Protocol (TCP). FTP was invented at M.I.T in 1971 before TCP/IP existed. The implementation of FTP is based on a client-server model and specifications are described by Request for Comment (RFC) document. This report presents the implementation of a file transfer applications based FTP. The application is built using python 3.6 run on Ubuntu 20.04.

II. BACKGROUND

FTP ensures data integrity and that files between the client and server are transferred without data loss. A client is defined as a computer that accesses the resources of another computer. A server is an always on host that responds to client requests. The interaction between the client and server is through commands defined by the RFC. To begin a FTP session two separate TCP connections are required; a data connection on port 20 and a command connection on port 21. FTP provides two ways in which a client can connect to a server namely active mode and passive mode.

A. Active Mode

In active mode the client connects to the server on port 21 using any unprivileged port ($P > 1024$) [1]. The

server then initiates a data connection between it and the client. After receiving the response the client obtains an unused unprivileged port to use for data connection. The problem with active mode is that most clients are behind a firewall which may not allow TCP connections on privileged ports.

B. Passive Mode

In passive mode the client is the one that initiates a TCP data connection unlike in active mode where its the server that does so. This modification solves the firewall problem in active mode since data connection flows from client to server. This means that data connection from the server uses an unprivileged port instead of port 20 [1].

C. Data formats

Different file formats are transmitted using specific data formats. The most commonly used data formats are ASCII and image type. All data transfers must be terminated with and end-of-file (EOF) [2], [3],

- The ASCII data format is used to transfer text files. The data is converted to 8 bits before and after transmission.
- The image type is used for efficient storage and retrieval of data in binary format. The data is converted into 8 bits before being transmitted.

There are mainly two transmission modes for FTP namely stream and block mode. In stream mode the data is sent as a stream of bytes. This mode allows any file type to be sent. Block mode allows a file to be transmitted in a series of blocks. The last block must indicate the end-of-file.

D. Problems with FTP

FTP was not designed with security in mind, the data is transmitted without encryption it is only encoded. This means that data can be intercepted through a packet analyzer. The protocol fails to handle high volume requests. Another issue with FTP is that it does not produce an error log file or report if there were problems during the transmission of a file. This means that it takes a while for issues to be resolved. An alternative to FTP is HTTP which is a browser based file transfer protocol [4].

III. SYSTEM OVERVIEW

The application is accessed through the terminal. The functionality of the application is based on the RFC 959, the response codes are also derived from this standard. The application was extensively tested on a single computer and error handling was implemented in order for the application not to crash during use.

A. Feature Overview

The basic core features are:

- Users can login after making requesting a connection to the server. The user logs in by providing a user name and password to gain access to the other features of the application.
- sending and downloading data
- Selecting transfer mode and data formats

Auxiliary features include:

- Printing current working directory
- changing working directory
- list directory files
- creating a new directory and deleting a file or directory.
- Multi threading is also used to allow multiple users to connect to the server

These features allow the user to manage their file structure in the server. Some functionality is not implemented.

B. Features not implemented

- All minimal client features were implemented as per the RFC.
- Server features not implemented include ability to interact with standard FTP client.
- The application has only been tested on a single machine.

IV. COMMAND OVERVIEW

The client and server were developed simultaneously. In order to a connection between server and client a TCP connection is initiated using sockets. The code below is used to declare in a TCP connection in the server file. This code specifies the that the application is using stream mode to transfer data, the socket is binded to TCP_IP which is "127.0.1.1" by default but can be changed. Accept is used to accept new connections, it returns a socket object and the client address.

In order to connect to the server the user must input the FTP command, if a connection is successfully established a 220 response code is returned to the client.

```
commandSocket =
    socket.socket(socket.AF_INET,
        socket.SOCK_STREAM)
commandSocket.bind((TCP_IP, TCP_PORT))
commandSocket.listen(1)
CONN, clientControlAddress =
    commandSocket.accept()
```

A. login: FTP

The login process is needed to verify a users details. User are given default usernames and passwords. The login in details are saves in a text file in the server. Only the administrator who has access to the log file in the server has writes to edit the document. This data is retrieved and stored in a dictionary in the code. This is to ensure that we have unique usernames and we can quickly search for a user in the log file. To login the user inputs the command USER followed by their user name. If the user name is exists in the database 331 is returned to the user whereby they are prompted to input a password using PASS with their password if the password is valid response code 230 is sent to the client telling them that they have successfully logged in. After a successful login the current working directory is set to the users folder. The code is structured in such a way that the user will only have access to the commands after logging in.

The transfer mode is set to active mode by default meaning that the server listens for a connection from a client and port 20 is used for data and port 21 for exchanging commands.

B. Print current working directory: PWD

In order to print the current working directory the user uses PWD. A response code of 257 is sent to the client to indicate the command was executed successfully. The implementation uses python os module. The os pwd command prints the working directory starting from the host machine home. This is an issue in this application because it causes security concerns. Users should not know the full path to their folder because it means that users know how the server is configures and which files. With such information users can backtrack and delete other users content or even the script used to run the server. To deal with this only the path starting from the current user's folder is displayed.

C. Make directory: MKD

This allows the user to create their own sub folders. This folder also uses os module to folder manipulation.

This is done using the MKD command. If the folder already exists response code 251 is returned to the client and if the new folder was successfully created 257 is sent to the client.

D. Change working directory: CWD

This command allows the user to access sub folders within their designated root folder. The command takes the pathname to the destination. Response code 250 is returned indicating that the file action is successful and 553 if it did not complete. The command will not complete if the pathname does not exist.

E. List files: LIST

The LIST command is used to list directory contents. The contents displayed are from the current working directory of the user in the server. A response code of 150 indicated that the command successfully executes.

F. Remove files: DELE

The DELE command is used to delete or remove a file within a directory or This command takes the file that the user wishes to delete as an argument. To delete a file we use `os.remove(file name)`.

G. Remove files: RMD

RMD is used to delete a directory and its contents. The command is very similar to the DELE command. The only difference is that it uses `shutil.rmtree(file name)` to delete a folder and its contents.

H. Uploading files to server: STOR

In order to upload files the client to server the user uses the STOR command to do. The command takes in the file name with file extension as an argument. The file the user intends on uploading to the server must be in the same directory as the client code. This is because the client code determines if the file exists before attempting to upload it. The response code 150 is used to indicate a successful file upload.

I. Downloading files from server: RETR

RETR is used to download previously stored files in the server. This command accepts the filename with file extension as an argument. A response code of 150 indicates a successful download from the server.

J. Quit connection: QUIT

The user can stop a TCP connection between themselves and the server by using the QUIT command. The command closes the previously opened TCP object that was initiated that was opened using the conn command.

K. Help: HELP

The help command displays available ftp commands to the user. The help command does not have a response code since it only implemented on the client side.

L. System: SYST

This command returns the underlying server operating system.

M. Passive mode: PASV

This command is used to initiate passive mode instead of the default active mode. This command tells the client to listen for a connection from the server.

N. Data port: PORT

This command is used to specify a data port for transmission. The commands takes a 32-bit host address and 16-bit port address. These are concatenated to be `h1,h2,h3,h4,p1,p2` where `h1` is the high order 8-bit representing the host address

V. FEATURE IMPLEMENTATION

This section provides a detailed code implementation of the implemented commands and the GUI used by the client.

Client side code function implementation is wrapped in try catch blocks in order avoid the application from crashing and for error handling, exceptions are thrown if the client code receives an error. All methods in the client side code start by sending the user command to the server. The methods in the server and client have the same name as the commands used.

The client methods send commands to the server received from the user through the GUI, server then calls the appropriate method. The struct module is used to interpret and send binary data across a network. The pack method is used to transmit the data format and data in a compact form. The unpack method is used to unbudle the transmitted data. Below we discuss some of the code implementation in detail which might be confusing to the reader the other methods are simple to understand and contain documentation.

A. RETR

A file is broken down and sent in parts defined by the buffer size. We use a for loop to send the pieces to the client in a stream of bits. The file is opened using 'rb' mode on the server side. The module read is used to convert the file contents into a string the file contents are sent as a string this is found between line 219 - 223 in the server code

The file is opened in 'wb' mode on the client side. This means that the file is opened in binary mode, the mode also ensure that the data remains as is during transmission. The receiving loop on the client is controlled by the file size to be received, this means that the received file size is less that the one in the server because of how the code is structured.

B. STOR

Likewise this command operates in a similar method like RETR, the difference is that operations are reversed. The server accepts the file in 'wb' and the client opens the file in 'rb'. The file is opened using read and is sent to the server as a string between line 148-143 in client code. The server receives the file in a while loop between line 101 - 105.

C. LIST

To list file we use `os.listdir(os.getcwd())` to get the list of files and sub directories in the current directory. Using this we get a file listing as strings. The list is then traversed using a for loop to send the file list to the client. The client receives the number os files in the directory first in order to bound the receiving for loop line 179 - 188 in client code.

D. PORT

Determining the data port is solely determined by the server code. The server converts client IP address into a port number. In order to do this we multiply the 5th octet(sequence of 8bits) and add it with the 6th octet. This is done is line 187 in server code. The new port is then used to unitize the socket object.

E. PASV

In order to start a passive mode connection the PASV command initiates a connection from the server to the client in order to avoid firewall issues from the client machine. In order to do this we randomly select unprivileged port between 1024 - 10000. To get the IP address from port we divide the and use modulo operator (divmomde in code) to get a pair of two numbers. Those are then used to represent the server IP address.

F. GUI

The GUI is implemented using PyQt5 which python wrapper for Qt v5. PyQt5 provides a framework for GUI development in python [5] . The GUI was developed after the application was tested in the terminal, including the GUI introduced new complexities. The GUI was designed using Qt designer. In order to link the client implementation and the GUI we import the functions from the client code implementation. This makes the code simpler to understand and maintain. The GUI is adjustable to meet the users needs.

G. Multithreading

Multithreading is used to allow multiple users to simultaneously connect to the same server. The maximum number users of allowed is limited to 5. To close all connections between users a for loop is used to join all threads from the users. The same client code is used for both threaded and serial code.

VI. IMPLEMENTATION CRITICAL ANALYSIS

The GUI is minimal and can be improved to include more functionality such as selecting a file from the clients machine nor just the current directory. The GUI can also be made more robust to include error handling at the moment the GUI freezes when the server receives asynchronously instructions. The instructions should be synchronised for the GUI to not crash. The STOR and RETR command can also be improved to allow the user the user to send or receive multiple file at a time.

The LIST command also facing a similar problem of file listings being out of sync. This results in unpredictable behavior in producing consistent results. The GUI code as also be converted to a more streamline installation process. This can be through converting the code to an executable or appropriate format to supported by the operating system.

The STOR and RETR commands are capable of handling text, images, and multimedia files and pdf files in active mode see figure fig1 - fig3 appendix. However it fails to handle document file such as pdf and word documents and multimedia files in passive see fig 4- fig 19 in appendix.

Another issue occurs when one does traffic monitoring on the application. When a user is changing a directory the full path is exposed when viewed in wireshark. This a security concern because the user can use this knowledge to access other directories in the server. In order to avoid this the server administrator must edit each directory property to only allow a specific access to the appropriate user. Directory traversal should also be limited to the

users parent directory. Common issues with TCP are also evident in this application. The data and commands are all sent in plain text, which means passwords and user names can be intercepted. These results can be seen in the appendix.

VII. CONCLUSION

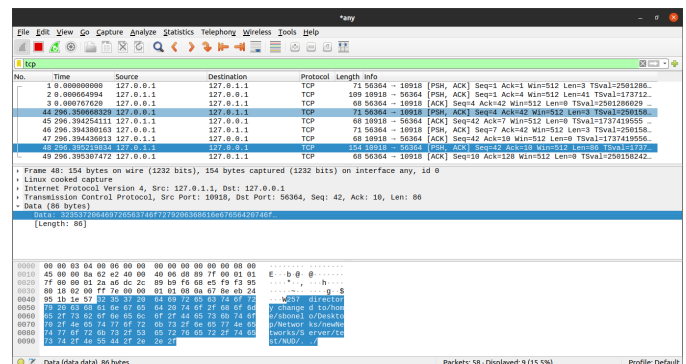
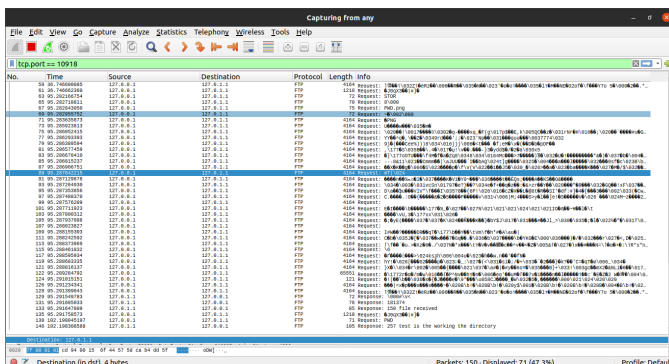
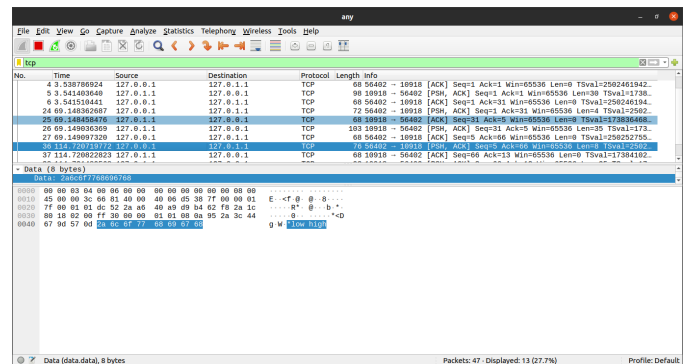
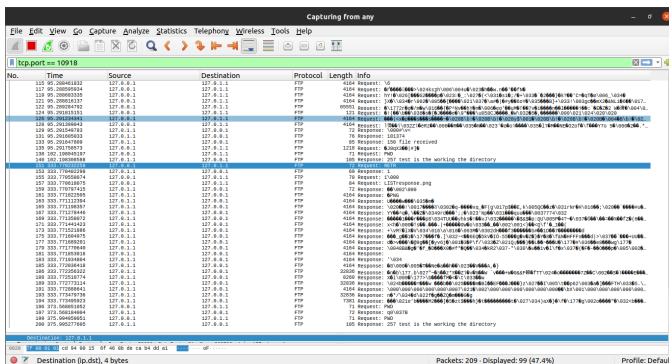
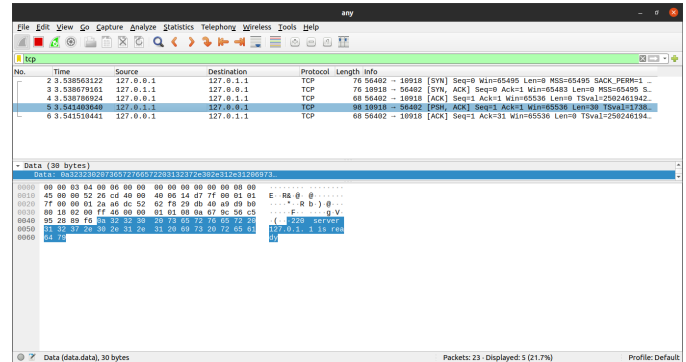
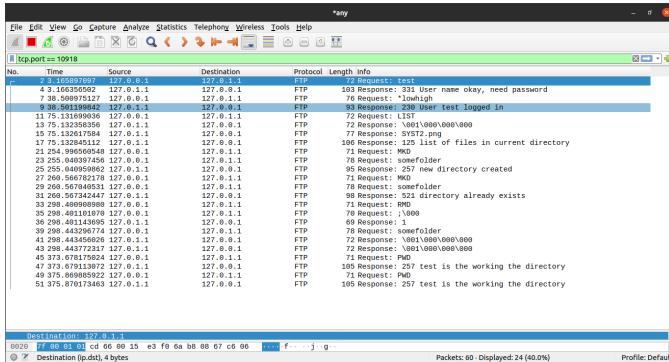
In this paper we presented an implementation of a file transfer application using sockets. The implementation is done using python. A user simple interface is provided for the user in order for them to have a better experience with the application. A multithreaded version of the server is also implemented. Wireshark is used to perform a packet interception on the application. The application can be further improved and made more robust to support more file formats.

REFERENCES

- [1] S. Site, "Active ftp vs. passive ftp, a definitive explanation." <https://slacksite.com/other/ftp.html>. Accessed: 2020-06-10.
- [2] W3, "File transfer functions." https://www.w3.org/Protocols/rfc959/4_FileTransfer.html. Accessed: 2020-05-30.
- [3] J.Postel and Reynold, "File transfer protocol (ftp)." <https://tools.ietf.org/html/rfc959>. Accessed: 2020-03-21.
- [4] A. Hughes, "The business problem of using file transfer protocol (ftp)." <https://www.cleo.com/blog/knowledge-base-file-transfer-protocol-business-problem>, note = Accessed: 2020-06-18.
- [5] Pypi, "Pyqt5." <https://pypi.org/project/PyQt5/>. Accessed: 2020-06-20.

APPENDIX

The results below are from traffic monitoring the application using wireshark. The screenshots contain the command and server response code.



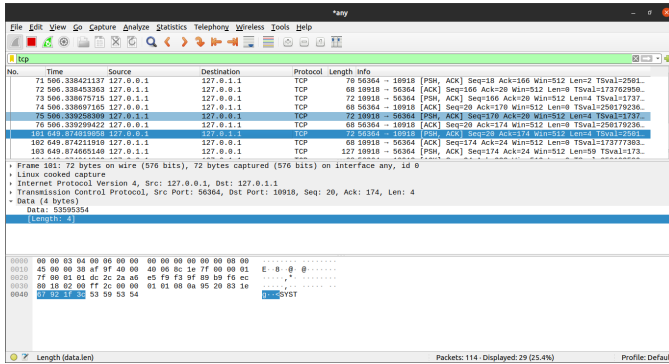


Fig. 7: SYST command

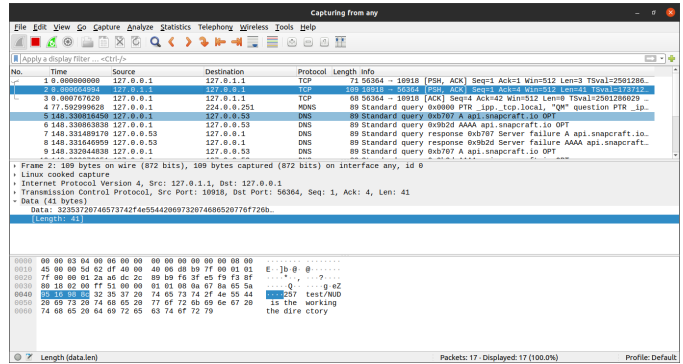


Fig. 10: SYST

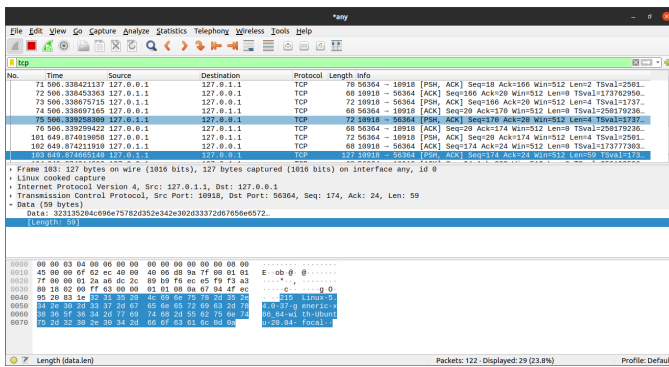


Fig. 8: SYST server response

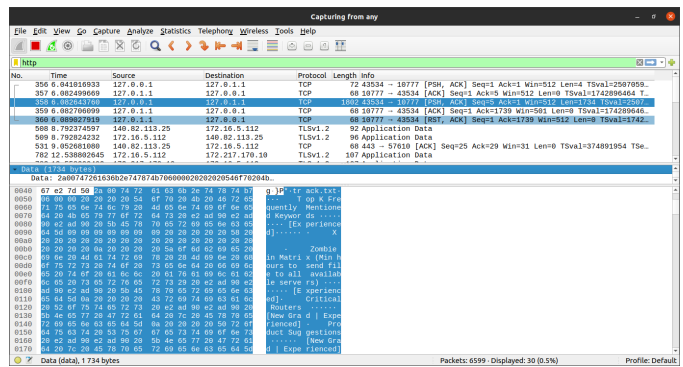


Fig. 11: STRO packet interception

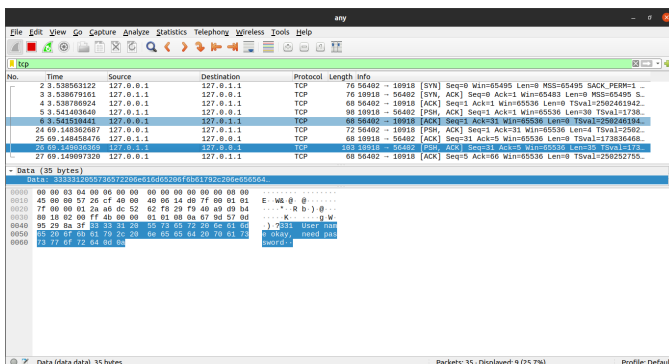


Fig. 9: user authorisation response

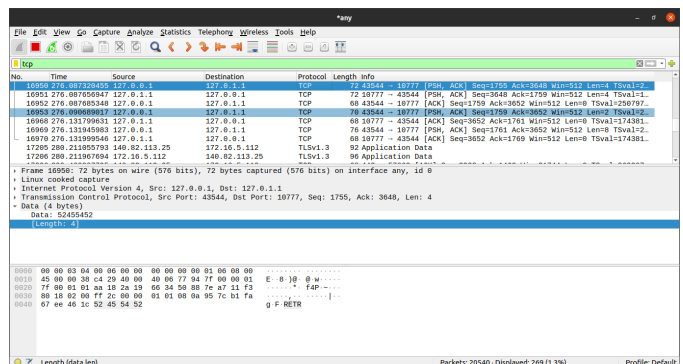
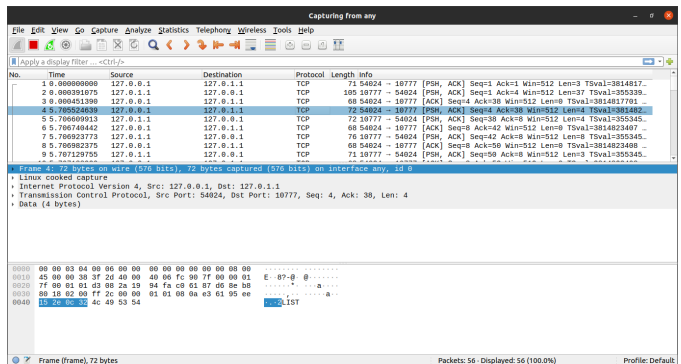
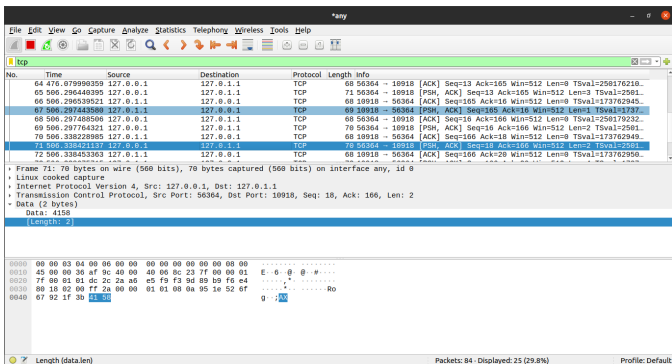
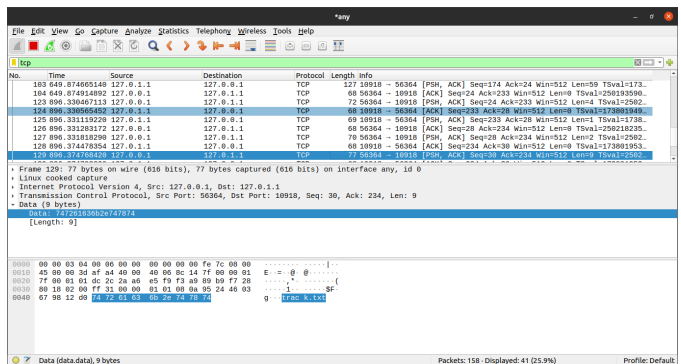
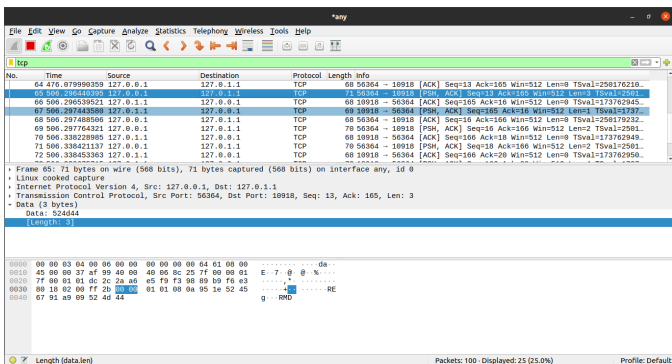
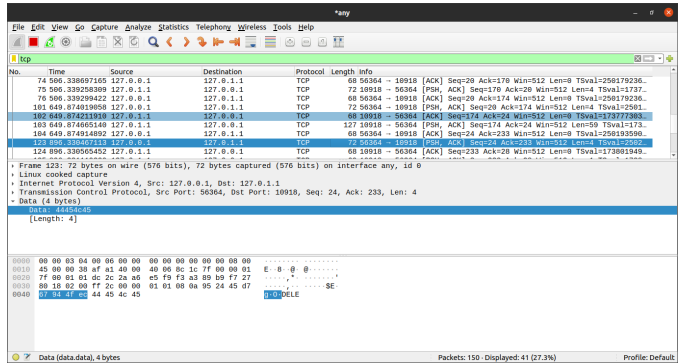


Fig. 12: RETR packet interception



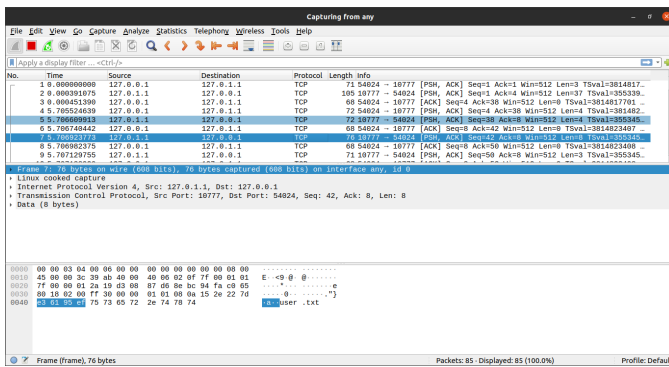


Fig. 19: LIST server response