COMP3310 Assignment 2 Report

U7493339

1. Introduction

1.1 Introduction

This report describes my Gopher client implemented using native socket programming in Python. The client was tasked with crawling the Gopher hierarchy hosted at `comp3310.ddns.net` on port `70`, classifying and indexing content, identifying the smallest and largest files, detecting malformed entries, and probing referenced external servers.

2. Implementation

2.1 Gopher Protocol

The client adheres to the Gopher protocol by sending requests in the form of:
<selector>\r\n
and reading responses until EOF. Each entry is parsed as per the standard:
<item_type><display_string>\t<selector>\t<host>\t<port>

2.2 Crawling Logic

The crawler:
- Maintains a set of visited selectors to prevent infinite loops.
- Parses directory (type 1) and recursively crawls internal paths.
- Collects text files (type 0) and binary files (type 9).
- Detects and logs malformed entries (type 3 or incorrect formatting).
- Attempts to connect to external servers and records their availability.

2.3 File Analysis

- Text file contents are read and stored.

The client identifies and prints:
- The smallest text file (by content length).
- The largest text file.
- The smallest and largest binary files (by byte length).

2.4 Fault Handling

- All failed or malformed requests are stored in an errors list.
- Timeouts and connection refusals are handled with exception blocks.
- Unknown or unsupported item types are skipped.

3. Output and operation

3.1 Output when running code

An example output when running my code is as follows:

```
[2025-04-23 22:31:37] Sending request:
[2025-04-23 22:31:37] Sending request: /rfc1436.txt
[2025-04-23 22:31:37] Sending request:
[2025-04-23 22:31:38] Sending request: /acme
Connection failed to comp3310.ddns.net:70 for selector '/acme/about' - timed out
[2025-04-23 22:31:43] Sending request: /acme/products
[2025-04-23 22:31:44] Sending request: /acme/products/anvils
[2025-04-23 22:31:44] Sending request: /acme/products/pianos
[2025-04-23 22:31:44] Sending request: /acme/products/paint
[2025-04-23 22:31:44] Sending request: /acme/products/traps
[2025-04-23 22:31:44] Sending request: /acme/contact
[2025-04-23 22:31:45] Sending request: /maze/17
[2025-04-23 22:31:45] Sending request: /maze/18
[2025-04-23 22:31:45] Sending request: /maze/19
[2025-04-23 22:31:45] Sending request: /maze/statuette
[2025-04-23 22:31:46] Sending request: /maze/floppy
[2025-04-23 22:31:46] Sending request: /maze/20
[2025-04-23 22:31:46] Sending request: /maze/21
Connection failed to comp3310.ddns.net:70 for selector '/maze/22' - timed out
[2025-04-23 22:31:51] Sending request: /misc
Connection failed to comp3310.ddns.net:70 for selector '/misc/empty' - timed out
[2025-04-23 22:31:56] Sending request: /misc/empty.txt
Connection failed to comp3310.ddns.net:70 for selector '/misc/nesta' - timed out
[2025-04-23 22:32:01] Sending request: /misc/nonexistent
[2025-04-23 22:32:01] Sending request: /misc/binary
[2025-04-23 22:32:01] Sending request: /misc/encabulator.jpeg
[2025-04-23 22:32:01] Sending request:
/misc/loooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooong
Connection failed to comp3310.ddns.net:72 for selector '' - [Errno 61] Connection refused
Connection failed to comp3310.ddns.net:73 for selector '' - timed out
Connection failed to comp3310.ddns.net:74 for selector '' - [Errno 61] Connection refused
[2025-04-23 22:32:06] Sending request: /misc/more
[2025-04-23 22:32:06] Sending request: /misc/malformed1
Malformed line: b'1Some menu - but on what host???\t/misc/malformed1/file\t'
[2025-04-23 22:32:07] Sending request: /misc/malformed2
Connection failed to comp3310.ddns.net:70 for selector '/misc/firehose' - timed out
[2025-04-23 22:32:12] Sending request: /misc/tarpit
Connection failed to comp3310.ddns.net:70 for selector '/misc/tarpit' - timed out
[2025-04-23 22:32:27] Sending request: /misc/godot
Connection failed to comp3310.ddns.net:70 for selector '/misc/godot' - timed out
[2025-04-23 22:32:32] Sending request:

--- Summary ---
Number of Gopher directories: 44

Number of text files: 10
 - /rfc1436.txt
 - /acme/products/anvils
 - /acme/products/pianos
 - /acme/products/paint
 - /acme/contact
 - /maze/statuette
 - /maze/floppy
 - /misc/empty.txt
 -
/misc/loooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooong
 - /misc/malformed2

Number of binary files: 2
 - /misc/binary
 - /misc/encabulator.jpeg

Contents of the smallest text file '/misc/empty.txt':
.


Size of the largest text file: 37396 bytes
Size of the smallest binary file: 253 bytes
Size of the largest binary file: 45584 bytes

Number of unique invalid references (errors): 6
 - ('invalid', 0, '')
 - ('comp3310.ddns.net', 70, '/maze/22')
 - ('comp3310.ddns.net', 70, '/misc/empty')
 - ('comp3310.ddns.net', 70, '/misc/nesta')
 - ('invalid', 0, '')
 - ('comp3310.ddns.net', 70, '/misc/malformed1')

External servers referenced:
 - gopher.floodgap.com:70 -> UP
 - comp3310.ddns.net:72 -> DOWN
 - comp3310.ddns.net:73 -> DOWN
 - comp3310.ddns.net:74 -> DOWN
 - comp3310.ddns.net:71 -> UP
```

3.2 Output summary

Using the above example, the crawler successfully traversed the Gopher hierarchy and produced the following statistics:

- Directories Found: 44
- Text Files Found: 10
  Example: /rfc1436.txt, /acme/products/paint, /maze/statuette
- Binary Files Found: 2
  Example: /misc/binary, /misc/encabulator.jpeg
- Smallest Text File: /misc/empty.txt (Content: .)
- Largest Text File Size: 37396 bytes
- Smallest Binary File Size: 253 bytes
- Largest Binary File Size: 45584 bytes
- Invalid References Logged: 6
  Example: ('comp3310.ddns.net', 70, '/maze/22')
- External Servers Referenced:
  - gopher.floodgap.com:70 – UP
  - comp3310.ddns.net on ports 71–74 – Mixed availability

3.3 Operation

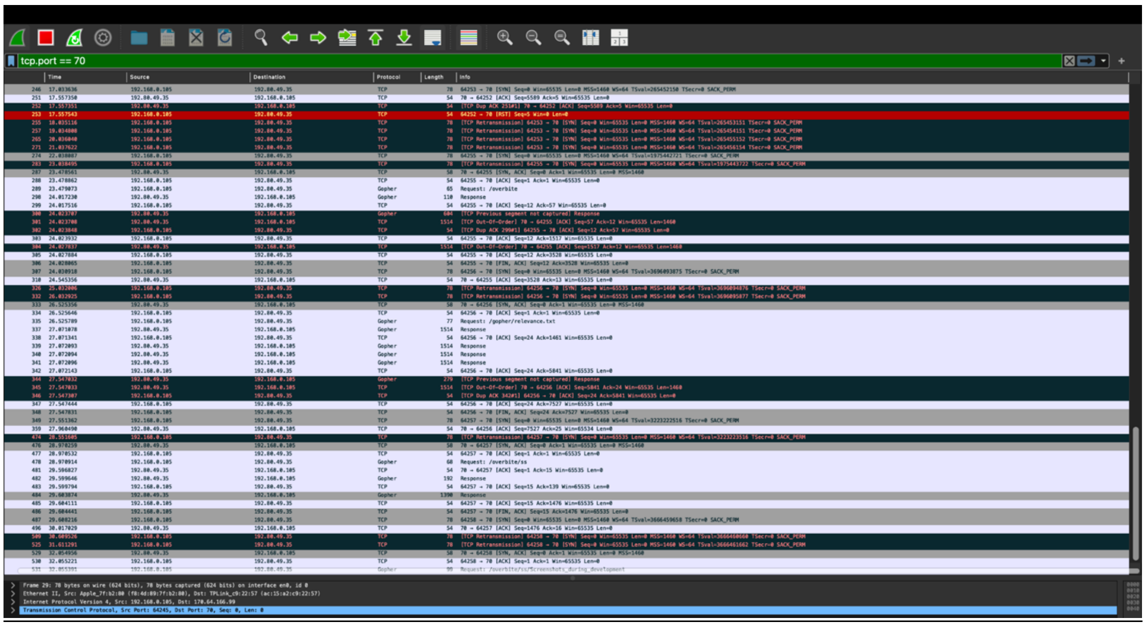
The code can be run by navigating to the correct directory (using cd) and then run 'python 3310task2.py'. The code the prompts the user to input the server host and server port, in this case 'comp3310.ddns.net' and '70' respectively. The code will then run, producing a similar output to what is shown above.

4. Challenges and Observations

- Several selectors timed out, indicating dead links or intentionally slow endpoints (e.g., /misc/tarpit).
- Malformed Gopher entries caused decoding failures, handled by error logging.
- Some entries reused the empty selector, likely due to incomplete formatting.
- Timestamped logs helped track request history and debug issues.

5. Wireshark

5.1 Wireshark screenshot



```
tcp.port == 70

     Time       Source          Destination     Protocol Length Info
246  17.033636  192.168.0.105   192.80.49.35    TCP      78     64253 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=265452150 TSecr=0 SACK_PERM
251  17.557350  192.80.49.35    192.168.0.105   TCP      54     70 → 64252 [ACK] Seq=5589 Ack=5 Win=65535 Len=0
252  17.557351  192.80.49.35    192.168.0.105   TCP      54     [TCP Dup ACK 251#1] 70 → 64252 [ACK] Seq=5589 Ack=5 Win=65535 Len=0
253  17.557543  192.168.0.105   192.80.49.35    TCP      54     64252 → 70 [RST] Seq=5 Win=0 Len=0
255  18.035116  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64253 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=265453151 TSecr=0 SACK_PERM
257  19.034808  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64253 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=265454151 TSecr=0 SACK_PERM
265  20.036040  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64253 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=265455152 TSecr=0 SACK_PERM
271  21.037622  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64253 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=265456154 TSecr=0 SACK_PERM
274  22.038087  192.168.0.105   192.80.49.35    TCP      78     64255 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=1975442721 TSecr=0 SACK_PERM
283  23.038495  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64255 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=1975443722 TSecr=0 SACK_PERM
287  23.478561  192.80.49.35    192.168.0.105   TCP      58     70 → 64255 [SYN, ACK] Seq=0 Ack=1 Win=65535 Len=0 MSS=1460
288  23.478862  192.168.0.105   192.80.49.35    TCP      54     64255 → 70 [ACK] Seq=1 Ack=1 Win=65535 Len=0
289  23.479073  192.168.0.105   192.80.49.35    Gopher   65     Request: /overbite
298  24.017230  192.80.49.35    192.168.0.105   Gopher   110    Response
299  24.017516  192.168.0.105   192.80.49.35    TCP      54     64255 → 70 [ACK] Seq=12 Ack=57 Win=65535 Len=0
300  24.023707  192.80.49.35    192.168.0.105   Gopher   604    [TCP Previous segment not captured] Response
301  24.023708  192.80.49.35    192.168.0.105   TCP      1514   [TCP Out-Of-Order] 70 → 64255 [ACK] Seq=57 Ack=12 Win=65535 Len=1460
302  24.023848  192.168.0.105   192.80.49.35    TCP      54     [TCP Dup ACK 299#1] 64255 → 70 [ACK] Seq=12 Ack=57 Win=65535 Len=0
303  24.023932  192.168.0.105   192.80.49.35    TCP      54     64255 → 70 [ACK] Seq=12 Ack=1517 Win=65535 Len=0
304  24.027837  192.80.49.35    192.168.0.105   TCP      1514   [TCP Out-Of-Order] 70 → 64255 [ACK] Seq=1517 Ack=12 Win=65535 Len=1460
305  24.027884  192.168.0.105   192.80.49.35    TCP      54     64255 → 70 [ACK] Seq=12 Ack=3528 Win=65535 Len=0
306  24.028065  192.168.0.105   192.80.49.35    TCP      54     64255 → 70 [FIN, ACK] Seq=12 Ack=3528 Win=65535 Len=0
307  24.030918  192.168.0.105   192.80.49.35    TCP      78     64256 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3696093875 TSecr=0 SACK_PERM
310  24.545356  192.80.49.35    192.168.0.105   TCP      54     70 → 64255 [ACK] Seq=3528 Ack=13 Win=65535 Len=0
326  25.032006  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64256 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3696094876 TSecr=0 SACK_PERM
332  26.032925  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64256 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3696095877 TSecr=0 SACK_PERM
333  26.525356  192.80.49.35    192.168.0.105   TCP      58     70 → 64256 [SYN, ACK] Seq=0 Ack=1 Win=65535 Len=0 MSS=1460
334  26.525646  192.168.0.105   192.80.49.35    TCP      54     64256 → 70 [ACK] Seq=1 Ack=1 Win=65535 Len=0
335  26.525789  192.168.0.105   192.80.49.35    Gopher   77     Request: /gopher/relevance.txt
337  27.071078  192.80.49.35    192.168.0.105   Gopher   1514   Response
338  27.071341  192.168.0.105   192.80.49.35    TCP      54     64256 → 70 [ACK] Seq=24 Ack=1461 Win=65535 Len=0
339  27.072093  192.80.49.35    192.168.0.105   Gopher   1514   Response
340  27.072094  192.80.49.35    192.168.0.105   Gopher   1514   Response
341  27.072096  192.80.49.35    192.168.0.105   Gopher   1514   Response
342  27.072143  192.168.0.105   192.80.49.35    TCP      54     64256 → 70 [ACK] Seq=24 Ack=5841 Win=65535 Len=0
344  27.547032  192.80.49.35    192.168.0.105   Gopher   279    [TCP Previous segment not captured] Response
345  27.547033  192.80.49.35    192.168.0.105   TCP      1514   [TCP Out-Of-Order] 70 → 64256 [ACK] Seq=5841 Ack=24 Win=65535 Len=1460
346  27.547307  192.168.0.105   192.80.49.35    TCP      54     [TCP Dup ACK 342#1] 64256 → 70 [ACK] Seq=24 Ack=5841 Win=65535 Len=0
347  27.547444  192.168.0.105   192.80.49.35    TCP      54     64256 → 70 [ACK] Seq=24 Ack=7527 Win=65535 Len=0
348  27.547831  192.168.0.105   192.80.49.35    TCP      54     64256 → 70 [FIN, ACK] Seq=24 Ack=7527 Win=65535 Len=0
349  27.551362  192.168.0.105   192.80.49.35    TCP      78     64257 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3223222516 TSecr=0 SACK_PERM
359  27.960490  192.80.49.35    192.168.0.105   TCP      54     70 → 64256 [ACK] Seq=7527 Ack=25 Win=65534 Len=0
474  28.551605  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64257 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3223223516 TSecr=0 SACK_PERM
476  28.970259  192.80.49.35    192.168.0.105   TCP      58     70 → 64257 [SYN, ACK] Seq=0 Ack=1 Win=65535 Len=0 MSS=1460
477  28.970532  192.168.0.105   192.80.49.35    TCP      54     64257 → 70 [ACK] Seq=1 Ack=1 Win=65535 Len=0
478  28.970914  192.168.0.105   192.80.49.35    Gopher   68     Request: /overbite/ss
481  29.596827  192.80.49.35    192.168.0.105   TCP      54     70 → 64257 [ACK] Seq=1 Ack=15 Win=65535 Len=0
482  29.599646  192.80.49.35    192.168.0.105   Gopher   192    Response
483  29.599794  192.168.0.105   192.80.49.35    TCP      54     64257 → 70 [ACK] Seq=15 Ack=139 Win=65535 Len=0
484  29.603874  192.80.49.35    192.168.0.105   Gopher   1390   Response
485  29.604111  192.168.0.105   192.80.49.35    TCP      54     64257 → 70 [ACK] Seq=15 Ack=1476 Win=65535 Len=0
486  29.604441  192.168.0.105   192.80.49.35    TCP      54     64257 → 70 [FIN, ACK] Seq=15 Ack=1476 Win=65535 Len=0
487  29.608216  192.168.0.105   192.80.49.35    TCP      78     64258 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3666459658 TSecr=0 SACK_PERM
496  30.017029  192.80.49.35    192.168.0.105   TCP      54     70 → 64257 [ACK] Seq=1476 Ack=16 Win=65535 Len=0
509  30.609526  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64258 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3666460660 TSecr=0 SACK_PERM
525  31.611291  192.168.0.105   192.80.49.35    TCP      78     [TCP Retransmission] 64258 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3666461662 TSecr=0 SACK_PERM
529  32.054956  192.80.49.35    192.168.0.105   TCP      58     70 → 64258 [SYN, ACK] Seq=0 Ack=1 Win=65535 Len=0 MSS=1460
530  32.055221  192.168.0.105   192.80.49.35    TCP      54     64258 → 70 [ACK] Seq=1 Ack=1 Win=65535 Len=0
531  32.055391  192.168.0.105   192.80.49.35    Gopher   99     Request: /overbite/ss/Screenshots_during_development

> Frame 29: 78 bytes on wire (624 bits), 78 bytes captured (624 bits) on interface en0, id 0
> Ethernet II, Src: Apple_7f:b2:80 (f8:4d:89:7f:b2:80), Dst: TPLink_c9:22:57 (ac:15:a2:c9:22:57)
> Internet Protocol Version 4, Src: 192.168.0.105, Dst: 170.64.166.99
> Transmission Control Protocol, Src Port: 64245, Dst Port: 70, Seq: 0, Len: 0
```

5.2 Wireshark Summary

Based on the Wireshark capture shown in the screenshot, the Gopher client generated a significant amount of TCP traffic over port 70 during its crawl of the server at comp3310.ddns.net. Multiple TCP connections were initiated from the client IP (192.168.0.105) to the server IP (192.168.0.35), with corresponding Gopher protocol requests and responses clearly visible. The presence of several TCP retransmissions and duplicate ACKs suggests instances of packet loss or delayed responses, which aligns with the timeouts and connection failures observed in the crawler's output, especially when accessing selectors like /acme/about and /misc/godot.

Additionally, the capture includes TCP reset (RST) packets, which indicate abrupt connection closures—likely from the server rejecting connections or from exhausted resources under load. This behaviour provides insight into the server's responsiveness and network reliability during the crawl. It also underscores the importance of the timeout and error-handling mechanisms implemented in the crawler code, which allowed the client to recover from failed or incomplete transactions without crashing or halting progress.