**Project Report #1: Email and Web Scraping with Python**

Daniel Hanna

Computer Networking (CSC 360)

The College of New Jersey

Department of Computer Science



**Part 0: Obtaining Familiarity with SMTP**

In attempting to understand how mail servers work, nslookup was used to determine which mail servers were used by TCNJ. It is clear from this command that there are two mail servers being used by TCNJ: mxgate1 and mxgate2. However, when attempting to utilize telnet in order to connect to either of these servers, the attempt times out, claiming that the resource is temporarily “unavailable.” However, this is expected, since as explained in the project outline, since the current computer’s host name does not match any registered host name on the DNS servers used by the college.

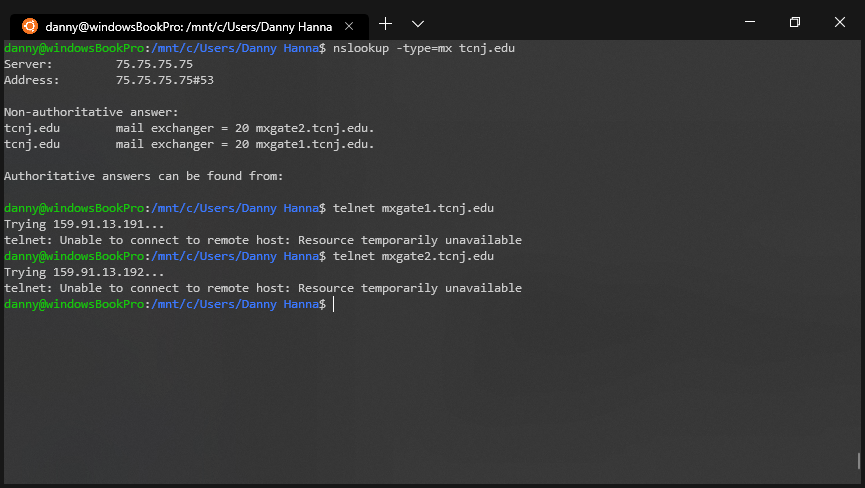


Figure 1: Using nslookup and telnet to access TCNJ’s mail servers

When attempting similar actions with Gmail’s servers, it can be seen that several mail servers exist, some of which are IMAP servers and some of which are SMTP servers. According to Google’s support documentation, Gmail’s SMTP servers use port number 587 for TLS (encrypted) protocol. Thus, openSSL must be utilized, as opposed to Telnet’s plaintext protocol.

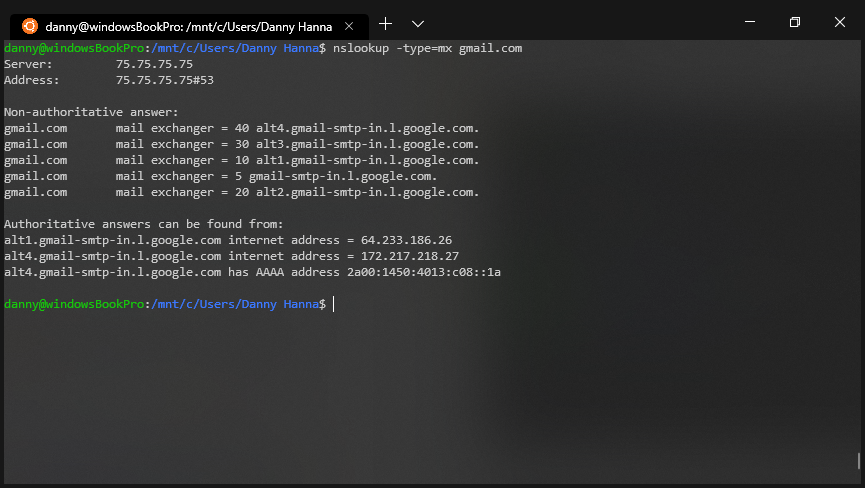


Figure 2: Using nslookup for Gmail's servers

In order to effectively connect to the SMTP mail server, a valid credential must be utilized. This can be done by Perl’s MIME, which encodes and decodes strings to and from a base64 encoding. This will provide an “authentication token” for connection. Figure 3 demonstrates the use of this command and the resulting output. Please note that some of the command has been covered for security.

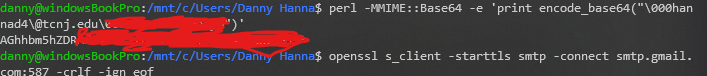


Figure 3: Perl MMIME to create a base64 encoding using login credentials

Connecting to the mail server and sending some mail then becomes a simple matter. OpenSSL is first used to connect to port 587 of the SMTP mail server.

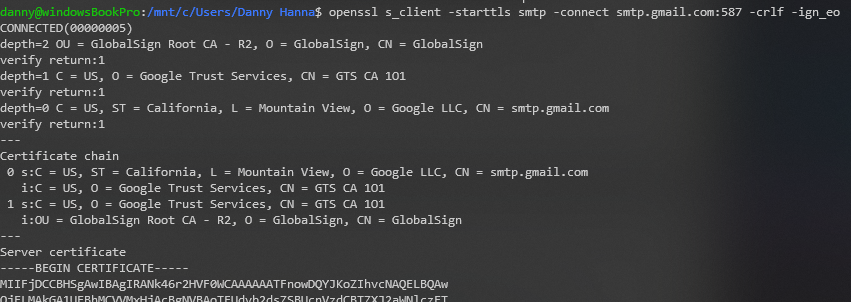


Figure 4: SSLing into server

Then authentication is performed by using the previously created base64 encoded sequence of characters. Note that for the purpose of security, the full character sequence is not included in the below figure.

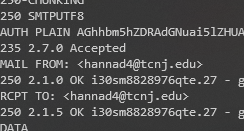


Figure 5: Authentication is Accepted

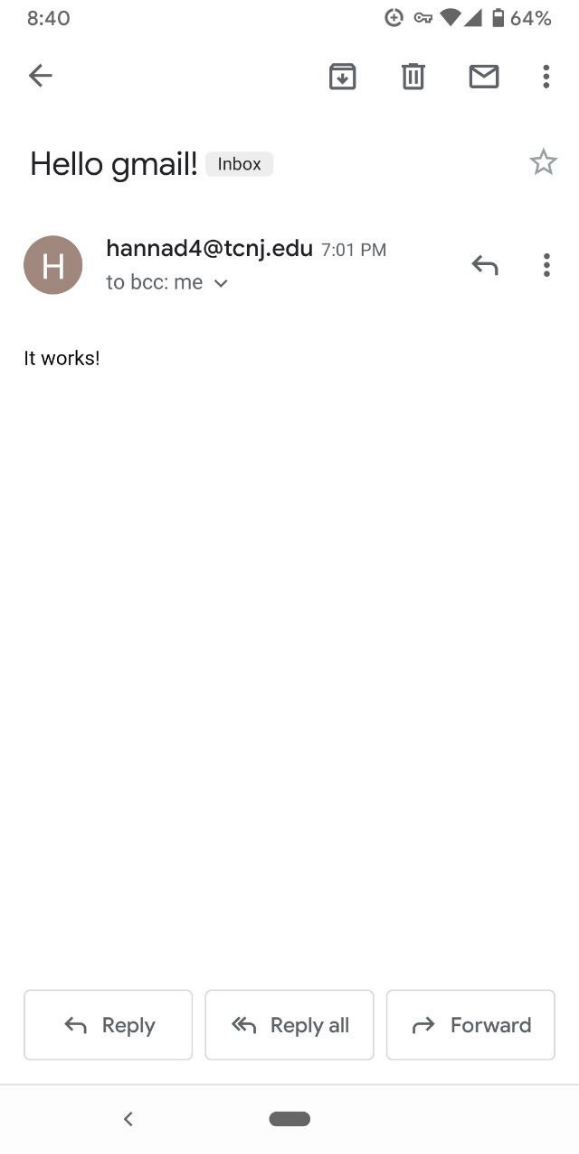
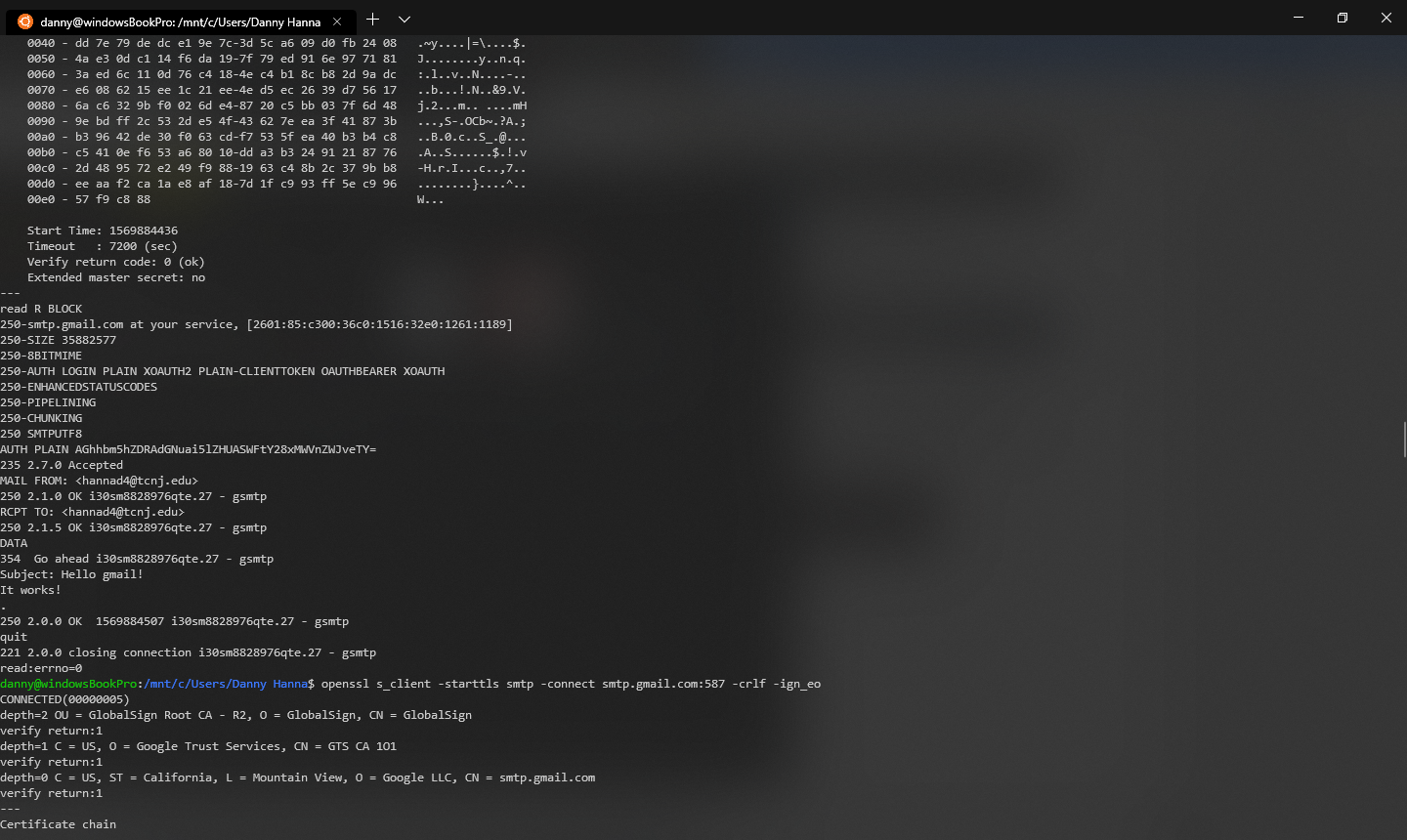
 It is then possible to send mail through the connection. Note the mail sent on the left half of the below figure, and the received mail on the right half.

Figure 6: Mail sent through SSL connection is received successfully

**Part 1: Extracting Stock Market Information**