UNIVERSITY OF WATERLOO

Faculty of Mathematics

Large Scale Mail Delivery System

Verizon Media (Yahoo!)

701 1st Ave, Sunnyvale, CA

Prepared by

Haomin Li

4B Computer Science

Student ID

April 12nd, 2020

MEMORANDUM OF SUBMITTAL

To: Yung Lin

From: Haomin Li

Date: April 12nd, 2020

Re: Large Scale Mail Delivery System

Dear supervisor,

This report, entitled "Large Scale Mail Delivery System", was prepared as my 4B term work report at Verizon Media (Yahoo!). This report is the third of the four work reports that the Co-operative Education Program requires for successful completion as part of my Computer Science Co-op degree requirements.

Throughout this internship, I have been working in the Yahoo Mail Delivery team mentored by Shashikiran and supervised by Yung. This report is written by me independently and has not received any previous academic credit. I would like to thank you for offering amazing assistance in the workplace whenever I need help, which I could use to develop this report.

Thank you for your time in assessing this report.

Eric (Haomin) Li

Table of Contents

Executive Summary

This report focuses on the large-scale mail delivery systems and the comparisons between different large-scale systems, such as PayPal and Yahoo Mail. After the 4-months experience of working in the Yahoo Mail Delivery team, I got my hands on one of the most popular email platforms, Yahoo Mail, and enhanced my understanding of different large-scale systems.

Email delivery is not a simple process. People in the world send billions of emails every day. Each email has to be sent to its recipient rapidly and accurately. There are many Mail delivery platforms, such as Yahoo Mail, Gmail, Outlook, and so on. How does a recipient from one mail delivery platform receive emails from another mail delivery platform? The process involves several components such as Mail User Agent, Mail Transfer Agent (MTA), Simple Mail Transfer Protocol (SMTP), and Message Pulling Protocols.

Besides the mail delivery system, the online payment system also plays a very important role in a large-scale delivery system. In this article, the differences between the online payment system and the mail delivery system would be discussed. PayPal would be taken as an online payment system compared with the mail delivery system.

Yahoo Mail is currently one of the most popular mail delivery systems. I will discuss the features and characteristics of Yahoo Mail that make it stand out in the industry.

1.0 Introduction

"With millions of emails sent every day, email has become a common mode of communication that is used by young and old. To have an email address – or even several – is something that is taken for granted; email has become as natural a communication channel as the telephone" (Christa, 2013). Email is essential to peoples' lives. It has been developed for many years since Ray Tomlinson created Email in 1971 (Menon, 2016). In 1971, Ray used @ sign to combine the user ID and server name together. By 1972, there were only hundreds of users communicated using emails. Currently, more than 600 million people around the world are using it for business or private usages. With many users and many emails are being sent, the accuracy and efficiency of sending emails need to be taken care of. Most people would not like their emails to be sent to the wrong destinations or their recipients receive the emails late. Some protocols and concepts are used to solve those problems.

Yahoo Mail is the first free webmail provider in the world, launched by Yahoo (now Verizon) in 1997. Now, it has over 320 million active users. According to Marketers technologies: Gmail vs Outlook vs Yahoo Mailbox, Yahoo Mail is mostly famous for its spam filtering technology (Ankit, 2019). Before the email is sent to the recipient, Yahoo Mail could classify it as spam or ham using its complicated classification algorithm. If email is classified as spam, it will either be blocked to its recipient or be placed in the recipient's spam box. If the email is classified as ham, it will be placed at recipients' inbox to read.

"Online payments began in the 1990s. The Stanford Federal Credit Union was the first institution to offer online banking services to customers in 1994" (A brief history of digital

payments, 2019). The early online payment system was not easy to use. It required users to understand data transfer protocols. In 1999, PayPal was invented together with the shopping platform, eBay. PayPal allows users to make online transactions without knowing the inside data transfer logic. It only requires users to click a button either on mobile apps or on the HTML page. Then, it will run some reverse Turing tests to determine if the transactions are a fraud. Up until now, PayPal has 286 million active users. (Millan, 2020)

2.0 Content

In the introduction, two problems of large-scale mail delivery system, accuracy as well as efficiency, has been discussed. However, we don't know how those problems were resolved. In the following paragraph, the structure of a large-scale mail delivery system will be explained. Besides, I will compare it with another large-scale distributed system: online payment system.

2.1 How does mail delivery system work

Assume Bob would like to send an email to Alice. Bob writes an email on a Mail User Agent (MUA). Then, the MUA sends the email with Simple Mail Transfer Protocol (SMTP) to the first Mail Transfer Agent (MTA). The message gets passed to multiple MTA and finally arrive at the last MTA called Mail Destination Agent (MDA). When Alice wants to check her email inbox. Alice's MUA pulls all the emails including the new one from MDA with a Message Pulling protocols. After that, Alice can read the email from Bob. (See figure 1.0)
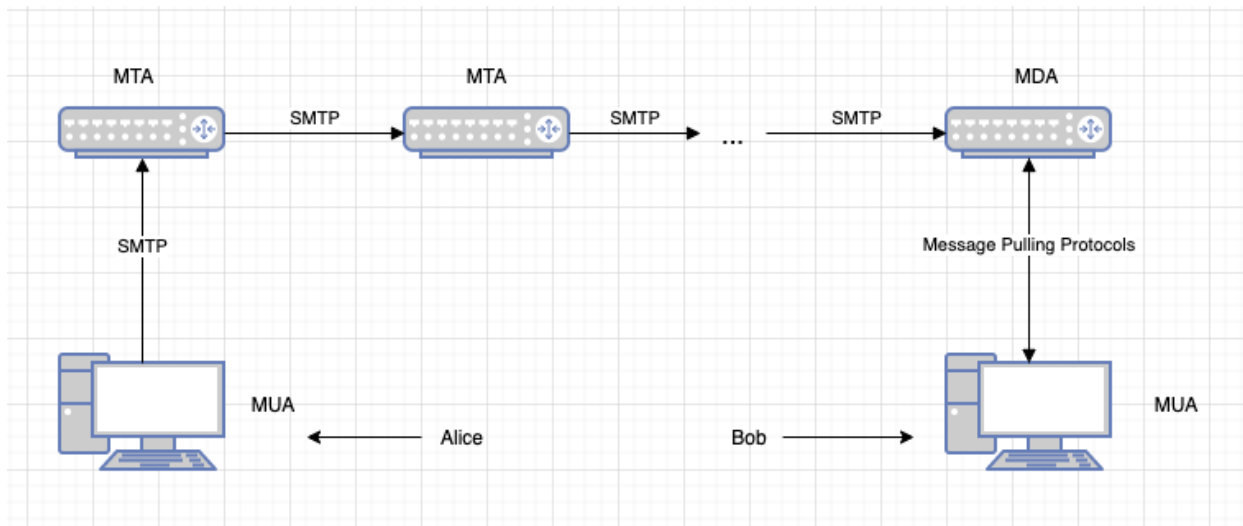
figure 1.0

2.1.1 Mail User Agent (MUA)

Mail User Agent, also called email client, is either a desktop application or mobile application for users to write emails or retrieve emails. For example, the Yahoo Mail application is an MUA. It has its brief and fresh UI design to improve usability.

2.1.2 Simple Mail Transfer Protocol (SMTP)

SMTP is the push protocol that is used to send emails to clients. While speaking in English, people need to follow specific grammar. Communication between two mail servers uses the same logic. SMTP is the 'grammar' that two servers use to communicate. (See figure 2)
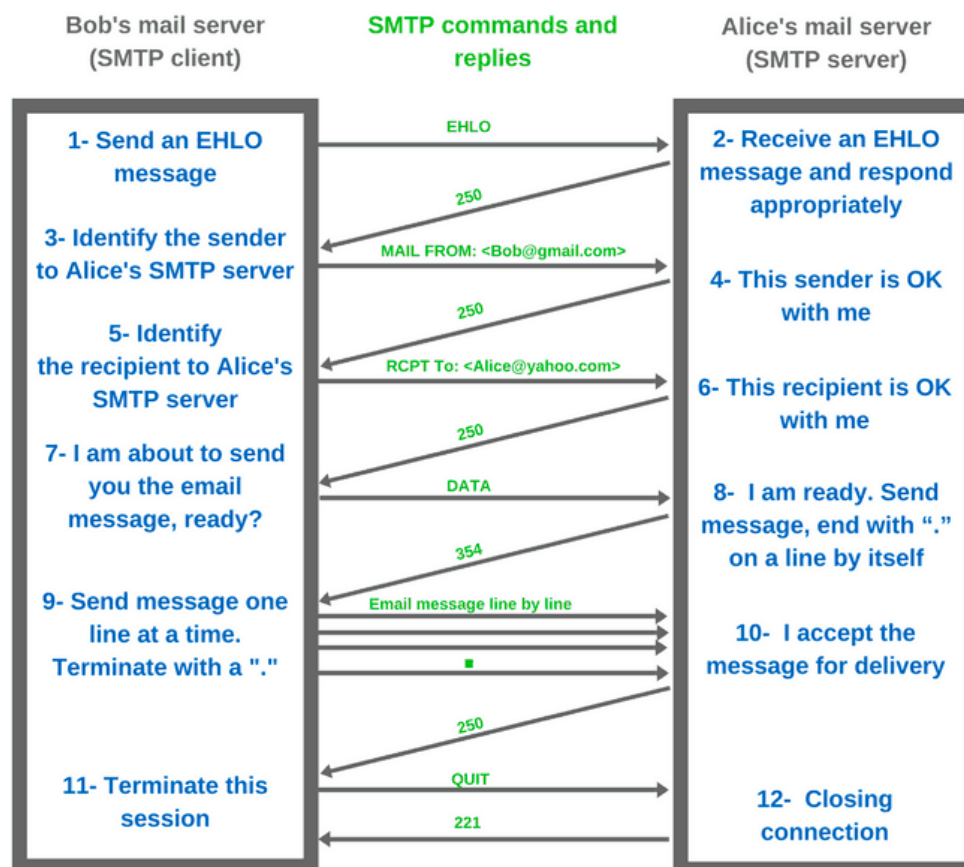
Bob's mail server (SMTP client) | SMTP commands and replies | Alice's mail server (SMTP server)

1- Send an EHLO message

EHLO →

2- Receive an EHLO message and respond appropriately

← 250

3- Identify the sender to Alice's SMTP server

MAIL FROM: <Bob@gmail.com> →

4- This sender is OK with me

← 250

5- Identify the recipient to Alice's SMTP server

RCPT To: <Alice@yahoo.com> →

6- This recipient is OK with me

← 250

7- I am about to send you the email message, ready?

DATA →

8- I am ready. Send message, end with "." on a line by itself

← 354

9- Send message one line at a time. Terminate with a "."

Email message line by line →
→
■ →

10- I accept the message for delivery

← 250

11- Terminate this session

QUIT →

← 221

12- Closing connection

figure 2 (SMTP protocol Explained, 2019)

2.1.3 Mail Transfer Agent (MTA)

A Mail Transfer Agent, also known as the mail server, is an element of the email delivery process. It receives emails from either another MTA or MUA. Then, it queries the MX records (a kind of DNS records) and selects another MTA to transfer emails. If the email fails to send to the next MTA, the current one will send an auto-response message back to the sender. Since there are too many emails being sent every second, MTA uses the store-and-forward to handle emails. The outgoing mails will be put in a queue repeatedly sending to the recipient MTA. The current MTA will not stop sending the email unless the recipient MTA confirms.

2.1.3 Message Pulling protocols

Different from SMTP, Message Pulling protocols are used by MUA to retrieve emails from servers. POP3 and IMAP are the two Message Pulling protocols in use. Take POP3 as an example. POP3 allows MUA to download emails from the remote users even if it is offline. While downloading emails, POP3 also removes the downloading emails in the server. In this way, POP3 can greatly reduce the space used in users' email servers. IMAP allows simultaneous access by multiple MUA. IMAP would be a better option than POP3 if users use multiple MUAs to manage their emails.

2.1.4 Why is Mail Delivery team important

Yahoo Mail has very large traffic. It has six to seven billion daily consumptions. For the Delivery team, there are two groups: Delivery and Antispam. It is well-known that mail delivery has the hardest work but earn the least appraisal from users. Most customers would focus on the UI changes of an MUA since the UI changes are very easy to discover in comparison to Mail delivery. The change may just be a slight change of number. For example, the delivery rate improves from 99.12% to 99.13%. Spam rate goes from 23% to 24%. However, those are actually the essence of email service. Since there are billions of consumptions every day, a 0.001% enhancement in delivery rate would result in more than millions of mails being successfully sent. A one percent improvement of the Spam rate would enable millions of spamming emails to be successfully identified. Most users would take emails being sent successfully for granted. If a few emails are not sent successfully, then users will be very angry

at the email service. The mail delivery team's works are crucial to email service but do not attract much of the public's attention.

2.2 Comparisons between mail delivery system and online payment system

The mail delivery system and the online payment system are both large-scale systems. Since both of them has very large consumptions, they are both implemented in an asynchronous way. In addition, they have many differences in terms of their requirement of accuracy and speed.

2.2.1 Asynchronous programming

Synchronous means that one can operate only operate one task at a time. Asynchronous means that one can operate multiple tasks at a time. Maybe it is hard to understand by just reading their definitions. For instance, assume mail delivery system is a restaurant, and sending emails is similar to ordering dishes. Customers come to a restaurant and wait for being seated at tables. Each waiter comes and guides customers to a table. The rest of the customers are still waiting for seats. Table A order three different dishes: dish one, dish two, and dish three.

In a synchronous way, each waiter only serves one table. The server for table A orders dish one and then carry it to the table A. After dish one is finished, the server order dish two and carries dish two to the table A, and so on. Note that the server has to wait for the specific customer to finish each dish. The server needs to wait for the kitchen to prepare each dish. There is plenty of idle space.

In an asynchronous way, each waiter can serve multiple customers. Firstly, a waiter places all the orders of table A to the kitchen. While waiting for the kitchen to prepare the dishes. The server can help other tables. The server will carry the cooked dishes to tables instead of only carry one each time.

If a restaurant has many customers waiting for seats, as we can see, an asynchronous way can make the best use of servers and is more efficient compared to the synchronous way. Thus, the mail delivery system and the online payment system are implemented in an asynchronous way. Note that most applications in real life are implemented in a synchronous way. The asynchronous system is very hard to implement.

2.2.2 Differences

The mail delivery system focuses more on speed, whereas the online payment system focuses more on accuracy. Users who use emails would expect to receive emails as soon as possible. If emails are failed to send, then users could send another one. However, in terms of the online payment systems, any mistake cannot be made regarding money. Thus, the online payment system would sacrifice the speed to ensure accuracy.

3.0 Conclusions

In conclusion, a large-scale mail delivery system has many complex components. A user writes an email on MUA. MUA sends the email via Simple Mail Transfer Protocol to MTA. The message goes through many MTAs and finally reaches the last MTA (also called MDA). The recipient on the other MUA pulls data from MDA using a message pulling protocol. To

implement the mail delivery system, asynchronous programming needs to be used because of the large consumptions. Besides, we also compared the mail delivery system with another large-scale system. The online payment system focuses more on accuracy while the mail delivery system pays more attention to speed.

Reference Page

1. Dürscheid, C., & Frehner, C. (2013). 2. Email communication. *Pragmatics of Computer-Mediated Communication*.

2. Menon, R. (2016, March 15). Ray Tomlinson's 'story' about inventing email is the biggest propaganda lie of modern tech history: Shiva Ayyadurai. Retrieved from https://economictimes.indiatimes.com/magazines/panache/ray-tomlinsons-story-about-inventing-email-is-the-biggest-propaganda-lie-of-modern-tech-history-shiva-ayyadurai/articleshow/51337493.cms

3. Prakash, A., Prakash, A., Prakash, A., Prakash, A., Prakash, A., Rodriguez, H., & Turner, M. (2019, November 2). Marketers Terminologes every marketer uses for marketing. Retrieved from https://aritic.com/blog/aritic-mail/marketers-terminologies-gmail-vs-outlook-vs-yahoo-mailbox/

4. Team, eP. (2019, January 31). A brief history of digital payments. Retrieved from https://blog.epayments.com/a-brief-history-of-digital-payments/

5. Urosevic, Urosevic, M., & Urosevic, M. (2020, January 28). 30 Incredible PayPal Statistics That Will Pay off in 2020. Retrieved from https://spendmenot.com/paypal-statistics/

6. Toppo, V., Vishal, Karim, Gabi, Gabi, Van, … KarimThank. (2019, August 19). SMTP protocol Explained (How Email works?). Retrieved from https://www.afternerd.com/blog/smtp/

Acknowledgments

I would like to thank my manager Yung and my mentor Shashi. Yung gave me this opportunity to work in his team and Shashi gave me plenty of mentorship in the project I was working on. They have been very helpful and patiently offered a lot of support since the first day I started working. They shared a lot of experience as senior engineers for the junior engineers working in the industry. I also want to thank the colleagues that I worked with. They gave me a lot of assistance during this internship.