Ana Paula Katsuda, A01025303
Andreína Sanánez, A01024927
Programación de estructuras de datos y archivos fundamentales
Prof. Jorge Rodríguez Ruíz
21 de Septiembre del 2021

Reporte Reto 1

1. Introducción

Recopilar información sobre las interacciones de los usuarios con distintos servicios, es de crucial importancia para identificar si existen accesos maliciosos en redes y proteger datos relevantes. Los accesos maliciosos tienen la posibilidad de afectar enormemente a empresas e individuos; puede costar la privacidad de éstos, e incluso puede tener impactos monetarios significativos.

Dada la importancia de dar seguimiento de dichas interacciones, el objetivo del presente escrito es utilizar algoritmos de búsqueda y ordenamiento para analizar un archivo que contiene las interacciones iniciales de usuarios en una empresa con servicios externos a ella. Lo anterior se realizará mediante la programación de vectores, de comparadores, la implementación de una clase Record y de los algoritmos mencionados anteriormente.

2. ADT

Nombre: Record
Datos:
- Fecha (string)
- Hora (string)
- IP Fuente (string)
- Puerto Fuente (int)
- Nombre Fuente (string)
- IP Destino (string)
- Puerto Destino (int)
- Nombre Destino (string)

Lista de operaciones:
- Método Record: guarda los datos extraídos del csv en sus respectivos atributos.
- Método Imprimir: muestra los datos del record.

3. Justificación de los Algoritmos

En cuanto a la implementación de la solución de esta sección del presente reto, se utilizaron los algoritmos Quicksort y Búsqueda Binaria, para realizar el ordenamiento y búsqueda respectivamente.

Comenzando con Quicksort es relevante mencionar que se eligió dicho algoritmo, ya que, a diferencia de otros algoritmos como Mergesort, tiene un buen balance entre sus complejidades temporal y espacial. Específicamente, Quicksort tiene una complejidad temporal promedio de $O(nlogn)$, en el peor de los casos siendo $O(n^2)$, y una complejidad espacial de $O(n)$. En contraste, a pesar de que Mergesort siempre tiene una complejidad temporal de $O(nlogn)$, sacrifica su complejidad espacial dado que es necesario hacer múltiples copias de los datos.

Por otro lado, se decidió utilizar búsqueda binaria puesto a que es el algoritmo de búsqueda más eficiente, teniendo una complejidad temporal de $O(log_2 n)$. Esto es

particularmente visible al compararlo con otros algoritmos como la búsqueda secuencial cuya complejidad temporal es $O(n)$.

4. **Resolución de las Preguntas**

a) ¿Cuántos registros tiene tu archivo?

Al utilizar la función *.size()* que determina el tamaño del vector que contiene cada uno de los registros, se obtuvo la siguiente cantidad:

```
Cuántos registros tiene su archivo?
Número de registros: 33251
```

Es importante comentar que para el manejo de los registros, se realizó una lectura del csv y se guardaron los registros en un vector.

b) ¿Cuántos récords hay del segundo día registrado? ¿Qué día es este?

Considerando que en el caso del csv proporcionado, los registros se encontraban inicialmente ordenados de menor a mayor fecha, fue posible obtener el segundo día al sumar una unidad al número de día del primer elemento del vector. Con esta información, se aplicó búsqueda binaria para encontrar un índice con dicho día, y así poder encontrar los límites superior e inferior de los registros correspondientes al día. Con lo anterior, se calculó la diferencia para obtener la cantidad de registros que se muestra:

```
Cuántos records hay del segundo día registrado? Qué día es este?
Día: 11
Número de records del segundo día: 3295
```

c) ¿Alguna de las computadoras pertenece a Jeffrey, Betty, Katherine, Scott, Benjamin, Samuel o Raymond?

Para conocer si alguna de las computadoras pertenecía a los individuos listados anteriormente, primero se ordenó el vector utilizando el orden gráfico léxico referente al nombre fuente. Posteriormente, se realizó la búsqueda individual para cada uno de los nombres, donde si el resultado es un número distinto a -1, significa que se encontró un índice con el nombre buscado. Cómo es posible observar a continuación, en el índice 29094 se encontró a un registro cuya computadora pertenece a el usuario Scott:

```
Alguna de las computadoras pertenece a Jeffrey, Betty, Katherine, Scott, Benjamin, Samuel, Raymond? (-1 = NO)
Resultado Búsqueda Jeffrey: -1
Resultado Búsqueda Betty: -1
Resultado Búsqueda Katherine: -1
Resultado Búsqueda Scott: 29094
Resultado Búsqueda Benjamín: -1
Resultado Búsqueda Samuel: -1
Resultado Búsqueda Raymond: -1
```

d) ¿Alguna computadora se llama *server.reto.com*?

Manteniendo el orden estipulado anteriormente, se volvió a realizar la búsqueda para el nombre fuente “server.reto.com” obteniendo que éste no se encuentra presente en el conjunto de datos registrados:

```
Alguna computadora se llama server.reto.com? (-1 = NO)
Resultado Búsqueda Server: -1
```

e) ¿Cuál es la dirección de la red interna de la compañía?

Al observar el csv, es posible notar que el IP fuente de los registros comienza con 172.24.238.X. A partir de lo anterior, se infiere que la dirección de la red interna de la compañía es 172.24.238.0, donde los últimos dígitos varían para cada usuario cuyo dispositivo está conectado a la red.

f) ¿Qué servicio de mail utilizan de todos estos: g*mail.com, outlook.com, protonmail.com, freemailserver.com*?

Siguiendo la línea de pensamiento planeada previamente, se ordenó el vector de registros a partir del nombre destino (esto también se hizo mediante el orden léxico gráfico). De manera similar a la búsqueda de nombre fuente, se realizó la búsqueda para cada uno de los nombres destino solicitados. En este caso, se encontró que el servicio de mail utilizado por los usuarios es protomail.com:

```
Qué servicio de email utilizan? (gmail.com, outlook.com, protonmail.com, freemailserver.com) (-1 = NO ENCONTRADO)
Resultado Búsqueda gmail: -1
Resultado Búsqueda outlook: -1
Resultado Búsqueda protomail: 24938
Resultado Búsqueda freemailserver: -1
```

g) Considerando solamente los puertos destino ¿Qué puertos abajo del 1000 se están usando? Lista los puertos e investiga qué aplicación/servicio lo utiliza generalmente.

Con el fin de obtener los puertos destino cuyo número es menor que 1000, se ordenaron los registros por puerto destino y se realizó la búsqueda de forma iterativa a partir del número 999 hasta hallar índice el primer número que cumple con la restricción. Una vez obtenido dicho número se puede iterar en el vector desde el índice establecido hasta el principio del vector, evaluando si ya se imprimió cierto número de puerto para no repetirlo y acceder al siguiente. Finalmente, se encontraron los siguientes puertos:

```
Considerando solamente los puertos destino, qué puertos abajo del 1000 se están usando?
993
465
443
135
80
67
53
```

A continuación, se explican los servicios que utilizan generalmente los puertos listados:

- **993:** lo utiliza el servicio IMAPS (protocolo de acceso a mensajes de internet). Un ejemplo de aplicación para este puerto son los servicios de email (Microsoft, s.f.).
- **465:** lo utiliza el servicio SMTP (Simple Mail Transfer Protocol). Es utilizado para enviar y recibir mensajes mediante correo electrónico. Se utiliza en conjunto con IMAP (IBM, s.f.).
- **443:** lo utiliza el servicio HTTPS (Hypertext Transfer Protocol Secure) y tiene la función de proteger la comunicación entre el usuario y el servidor (ClickSSL, 2021).
- **135:** lo utiliza el servicio RPC (Remote Procedure Call) para aplicaciones que requieren comunicación entre el cliente y el servidor (McNab, C., 2004).
- **80:** lo utiliza el servicio HTTP (Hypertext Transfer Protocol) y es utilizado para establecer comunicación con un servidor o con otra computadora en el internet (Fox, P, s.f.).
- **67:** lo utiliza el servicio DHCP (Dynamic Host Configuration Protocol) el cual le envía al puerto 67 los mensajes que el cliente le manda al servidor (Palo Alto Networks, 2021).
- **53**: lo utiliza el servicio DNS (Domain Name System) para las transferencias de zona (Howtouselinux.com, 2021).

5. Aportaciones

En cuanto a las aportaciones del equipo, es importante mencionar que el análisis de la situación problema, así como su resolución fue hecha en conjunto por las dos integrantes del equipo. Si bien ciertas partes fueron realizadas de manera individual, se tuvo una sesión de revisión para verificar el correcto funcionamiento de la incorporación de cada una de las aportaciones. A continuación, se mencionan las divisiones de tareas realizadas por cada integrante:

En conjunto:

- Código para la lectura y guardado de los registros
- Elaboración escrita del ATD
- Elaboración del reporte

Andreína Sanánez / A01024927

- Programación de los comparadores individuales para cada atributo (comparación entre valor y atributo del record/registro).
- Programación y adecuación de la clase Quicksort.
- Programación de la solución de las preguntas 1-3.
- Comentarios del programa.

Ana Paula Katsuda / A01025303

- Programación de los comparadores con dos parámetros (comparación de dos récords/registros)
- Programación de la adecuación para búsqueda binaria.
- Programación de la solución para los incisos 5-7.

6. Referencias

ClickSSL. (2021). Port 443 - what you need to know about HTTPS 443. [Sitio Web]. Recuperado de: https://www.clickssl.net/blog/port-443

Descubridor de puertos online. (s.f.).[Sitio web]. Recuperado de: https://es.adminsub.net/tcp-udp-port-finder

DHCP Overview. (2021). [Sitio Web]. Recuperado de: https://docs.paloaltonetworks.com/pan-os/8-1/pan-os-admin/networking/dhcp/dhcp-overview.html

Fox, P. (s.f.) Hypertext Transfer Protocol. Khan Academy. [Sitio Web]. Recuperado de: https://www.khanacademy.org/computing/computers-and-internet/xcae6f4a7ff015e7d:the-internet/xcae6f4a7ff015e7d:web-protocols/a/hypertext-transfer-protocol-http

IBM. (s.f.) Protocolo simple de transferencia de correo. [Sitio Web]. Recuperado de: https://www.ibm.com/docs/es/i/7.3?topic=information-smtp

Microsoft. (s.f.) ¿Qué son IMAP y POP? [Sitio web]. Recuperado de: https://support.microsoft.com/es-es/office/-qu%C3%A9-son-imap-y-pop-ca2c5799-49f9-4079-aefe-ddca85d5b1c9

McNab, C. (2004). Assessing Windows Networking Services. [Sitio Web]. Recuperado de: https://www.oreilly.com/library/view/network-security-assessment/059600611X/ch09.html

Understanding DNS port 5 with examples. (2021). [Sitio Web]. Recuperado de: https://www.howtouselinux.com/post/dns-port