# C++ Data structures to understand botnet

Authors: José Luis Madrigal Sánchez, Jorge Isidro Blanco Martínez and Alan Josué Melgar Fuentes

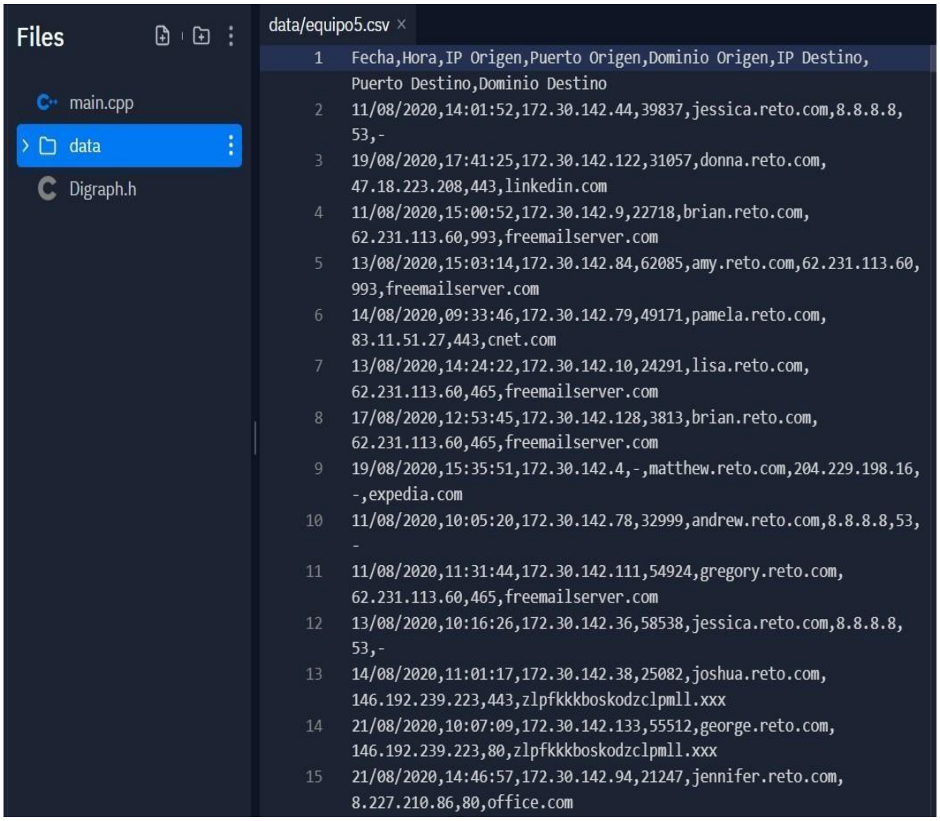
The structures which we worked with were vectors, linked lists, trees,graphs and hash tables. During the semester, we used a csv file that contained many registersof connections made by hosts in a network, including things like date, time, domains, IP addresses and ports.

Basically, we made scripts to sort the registers by different criteria and organized them in certain data structure so we can find some pattern or suspicious activity.

We answered 3 basic questions:

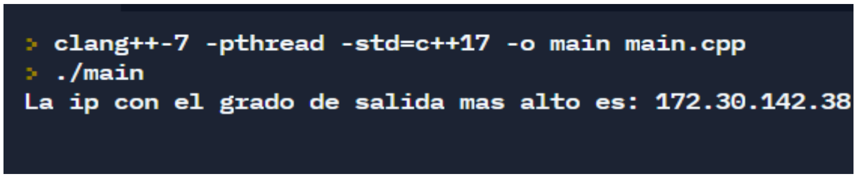
- Boot master (trees): 62.231.113.60
- First infected host (graphs): 172.30.142.38
- More attacked domain (hash tables): forbes.com

Here we can see the file that was used and was given by our teacher.



```
Fecha,Hora,IP Origen,Puerto Origen,Dominio Origen,IP Destino,Puerto Destino,Dominio Destino
11/08/2020,14:01:52,172.30.142.44,39837,jessica.reto.com,8.8.8.8,53,-
19/08/2020,17:41:25,172.30.142.122,31057,donna.reto.com,47.18.223.208,443,linkedin.com
11/08/2020,15:00:52,172.30.142.9,22718,brian.reto.com,62.231.113.60,993,freemailserver.com
13/08/2020,15:03:14,172.30.142.84,62085,amy.reto.com,62.231.113.60,993,freemailserver.com
14/08/2020,09:33:46,172.30.142.79,49171,pamela.reto.com,83.11.51.27,443,cnet.com
13/08/2020,14:24:22,172.30.142.10,24291,lisa.reto.com,62.231.113.60,465,freemailserver.com
17/08/2020,12:53:45,172.30.142.128,3813,brian.reto.com,62.231.113.60,465,freemailserver.com
19/08/2020,15:35:51,172.30.142.4,-,matthew.reto.com,204.229.198.16,-,expedia.com
11/08/2020,10:05:20,172.30.142.78,32999,andrew.reto.com,8.8.8.8,53,-
11/08/2020,11:31:44,172.30.142.111,54924,gregory.reto.com,62.231.113.60,465,freemailserver.com
13/08/2020,10:16:26,172.30.142.36,58538,jessica.reto.com,8.8.8.8,53,-
14/08/2020,11:01:17,172.30.142.38,25082,joshua.reto.com,146.192.239.223,443,zlpfkkkboskodzclpmll.xxx
21/08/2020,10:07:09,172.30.142.133,55512,george.reto.com,146.192.239.223,80,zlpfkkkboskodzclpmll.xxx
21/08/2020,14:46:57,172.30.142.94,21247,jennifer.reto.com,8.227.210.86,80,office.com
```

Finally, we can see the message that is generated after making all the necessary operations and filters, in this case, with the implementation of a graph, we got the node (IP address) with the highest output rate, which is the first infected host.



```
> clang++-7 -pthread -std=c++17 -o main main.cpp
> ./main
La ip con el grado de salida mas alto es: 172.30.142.38
```