Packet Capture

Cirillo Antonio - Di Pierno Andrea - Sabato Vincenzo

1. Definizione degli obiettivi

1.1. Cattura del traffico di rete ad intervalli di tempo crescenti al fine di analizzare le relazioni esistenti tra tempo di cattura, dimensioni dei file, numero di richieste a siti terzi e pacchetti Https.

2. Wireshark

2.1. WireShark è un tool multipiattaforma che consente di intercettare il traffico dati che transita su una specifica interfaccia di rete, fornendo informazioni riguardanti lo scambio di pacchetti dati. Nell'ambito di questo esperimento, Wireshark è stato utilizzato per la cattura dei pacchetti al fine di analizzarne le informazioni su sorgente e destinazione e l'utilizzo o meno del protocollo Https.

3. Beautiful Soup

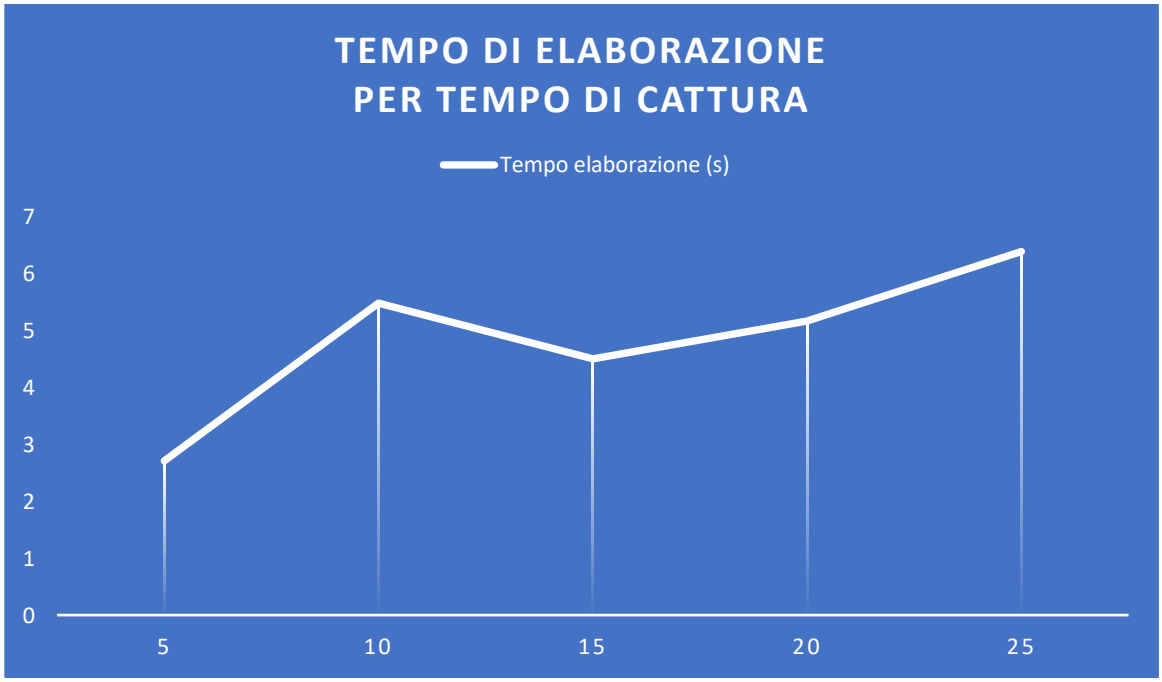
3.1. Beautiful Soup è una libreria compatibile con Python 2.7 e 3.8 che consente di elaborare in modo semplice le informazioni contenute in file HTML e XML. In questo esperimento, Beautiful Soup ha consentito di recuperare rapidamente l'elenco dei siti da visitare.

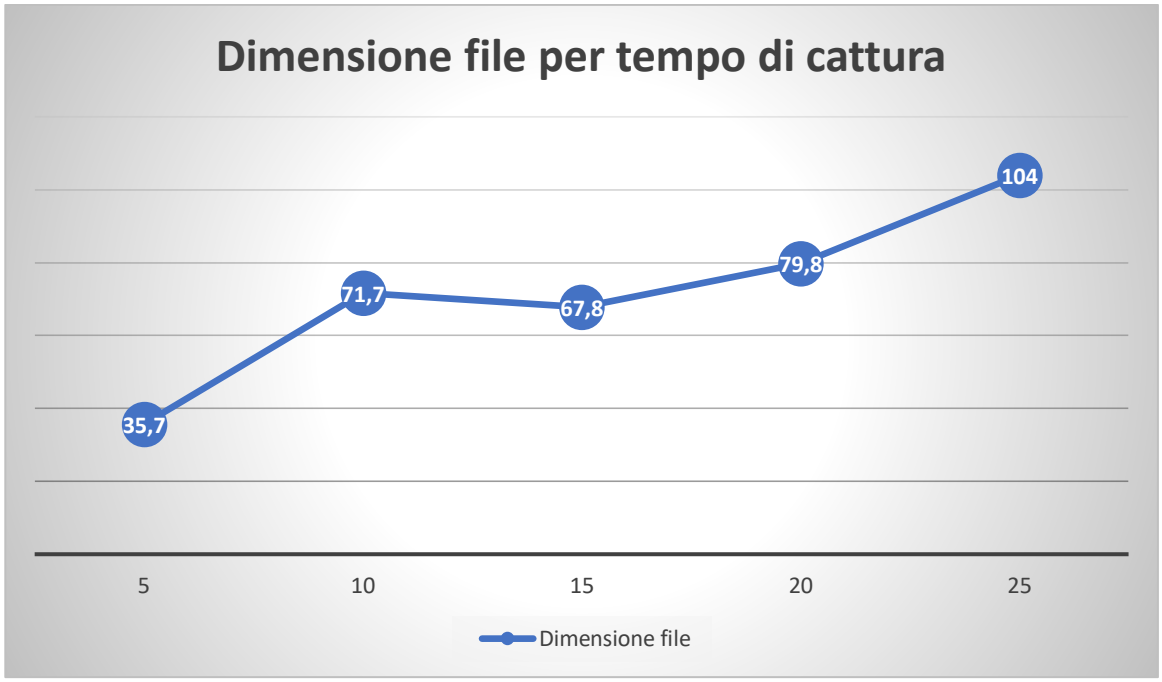
4. Descrizione dell'esperimento

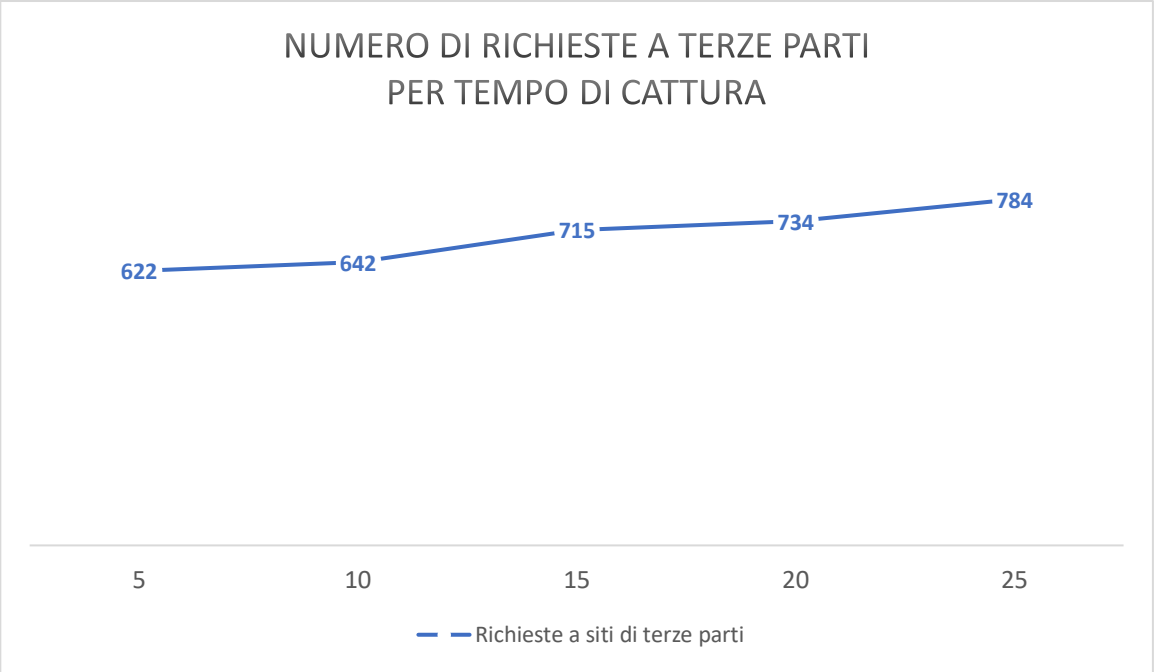
4.1. L'esperimento consiste nella cattura e nell'analisi dei pacchetti dati provenienti dalle richieste effettuate verso i 50 siti più popolari secondo Alexa, per intervalli di tempo di 5, 10, 15, 20 e 25 minuti di navigazione. Grazie a Beautiful Soup vengono recuperati gli indirizzi web dei siti da visitare, dei quali si ottiene successivamente l'indirizzo IP, inserito in un apposito Array. Una volta concluso il processo di recupero degli indirizzi IP, vengono effettuate le richieste ai siti web ed il traffico viene catturato tramite Wireshark.

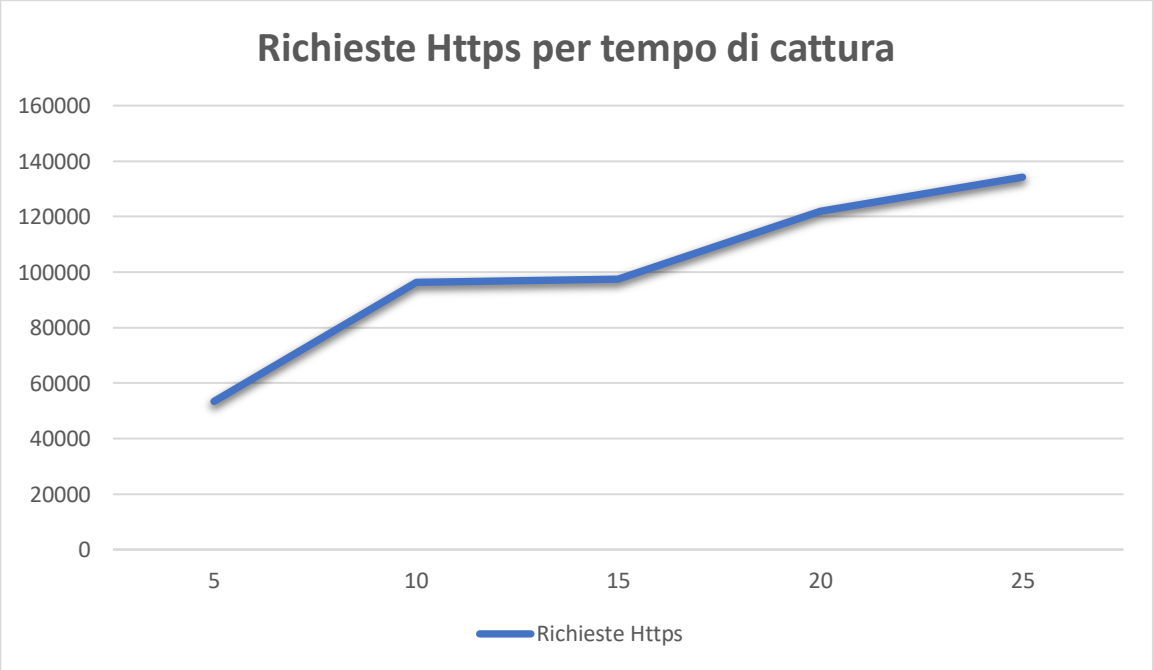
A cattura ultimata, i file ottenuti vengono analizzati, confrontando gli indirizzi IP dei vari pacchetti con l'indirizzo sorgente (quello dell'interfaccia di rete) e i 50 indirizzi IP associati ai siti web, al fine di verificare il numero di richieste a siti di terze parti effettuate. Inoltre, viene verificato per ciascun pacchetto l'utilizzo o meno del protocollo Https. Infine, ad analisi conclusa, viene fornito anche il tempo di elaborazione per ciascun file.

5. Risultati ottenuti









6. Conclusioni

Come si può evincere dai risultati in forma grafica, il tempo di cattura incide notevolmente sulla dimensione del file di log e sul numero di richieste Https effettuate. Il numero di richieste a siti di terze parti, invece, cresce lievemente con l'aumentare del tempo di cattura, per via dei refresh effettuati dalle risorse di alcune pagine (prevalentemente banner pubblicitari). Infine, il tempo di elaborazione aumenta lievemente a seconda delle dimensioni del file di log.