Secure Compression and Pattern Matching Based on Burrows-Wheeler Transform

Raffaele Ceruso Giovanni Leo

November 24, 2018

Idea

- Le compressed data structure permettono di creare una indicizzazione di grandi dataset in maniera efficiente. Tali strutture dati dovranno essere sicuramente memorizzate in qualche server di terze parti come ad esempio il cloud, questo per porta sicuramente problemi di privacy.
- L idea quindi quella di costruire una variante di tali strutture pi sicura, basata sulla trasformata di Burrows-Wheeler, e che sia in grado anche di eseguire il pattern matching.

Introduzione - 1

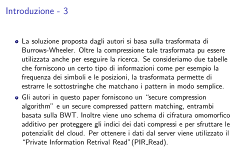
- La compressione dati non solo riduce lo spazio occupato dai file ma serve anche a migliorare la velocit di trasmissione in alcuni protocolli.
- Essendo in era in cui i dati sono diventati veramente importanti di conseguenza anche la compressione di tali dati diventata sempre pi necessaria.
- Gli autori del paper si sono focalizzati sullutilizzo di algoritmi di compressione, i quali non solo supportano una compressione di tipo lossless ma che garantiscono anche sicurezza.
- Il pattern matching una operazione fondamentale nel processing delle stringhe che permette di trovare tutte le occorrenze di un dato pattern in un dato testo.

Introduzione - 2

- Il pattern matching si vuole applicare anche ai file compressi
- Una possibile strategia potrebbe essere quella di decomprimere il file e cercare sul file decompresso ma tale strategia poco efficiente.
- Un approccio migliore sarebbe quello di ricercare direttamente sul file compresso
- Per fare ci abbiamo bisogno di indici di dati compressi i quali vengono salvati in server di terze parti causando problemi di privacy.
- Una semplice soluzione potrebbe essere quella di di comprimere e poi cifrare i dati ma tale soluzione stata dimostrata non tanto sicura.
- Una altra soluzione potrebbe essere quella di integrare compressione e cifratura in un unico passo ma anche questa soluzione presenta problemi di sicurezza.p_h

Introduzione - 3

- La soluzione proposta dagli autori si basa sulla trasformata di Burrows-Wheeler. Oltre la compressione tale trasformata pu essere utilizzata anche per eseguire la ricerca. Se consideriamo due tabelle che forniscono un certo tipo di informazioni come per esempio la frequenza dei simboli e le posizioni, la trasformata permette di estrarre le sottostringhe che matchano i pattern in modo semplice.
- Gli autori in questo paper forniscono un "secure compression algorithm" e un secure compressed pattern matching, entrambi basata sulla BWT. Inoltre viene uno schema di cifratura omomorfico additivo per proteggere gli indici dei dati compressi e per sfruttare le potenzialit del cloud. Per ottenere i dati dal server viene utilizzato il "Private Information Retrival Read"(PIR_Read).



Introduzione - 3

- La soluzione proposta dagli autori si basa sulla trasformata di Burrows-Wheeler. Oltre la compressione tale trasformata pu essere utilizzata anche per eseguire la ricerca. Se consideriamo due tabelle che forniscono un certo tipo di informazioni come per esempio la frequenza dei simboli e le posizioni, la trasformata permette di estrarre le sottostringhe che matchano i pattern in modo semplice.
- Gli autori in questo paper forniscono un "secure compression algorithm" e un secure compressed pattern matching, entrambi basata sulla BWT. Inoltre viene uno schema di cifratura omomorfico additivo per proteggere gli indici dei dati compressi e per sfruttare le potenzialit del cloud. Per ottenere i dati dal server viene utilizzato il "Private Information Retrival Read"(PIR_Read).

- La crittografia omomorfica una forma di crittografia che consente il calcolo su testi cifrati(In questo caso la somma), generando un risultato crittografato che, decodificato, corrisponde al risultato delle operazioni come se fossero state eseguite sul testo in chiaro.
- PIR_Read: un modo privato per ottenere dati dal server