Filtres de Bloom i Cryptohashing

Pol Casasayas Bryan Game Erik A. Kvam

23 d'abril de 2017

Resum

En aquest document s'intenta analitzar la diferència en la probabilitat de fals positiu d'un filtre de Bloom usant k funcions de hash directament o encriptant les claus prèviament amb SHA256. Per a tal efecte, primer s'intenta determinar quina família de hash dona un nombre més reduït de falsos positius, amb o sense rolling hash aplicat a sobre, i seguidaments es compara amb aplicar-li l'SHA256 a sobre.

1 Introducció

L'objectiu de la pràctica és, principalment, trobar quina és la millora en el número de falsos positius d'un filtre de Bloom fent servir o no SHA256 abans d'aplicar-lo. Per a tal efecte, s'ha dissenyat un experiment, que s'explica a l'apartat 2.3, que posa k funcions de hash envers les mateixes funcions amb SHA256 aplicats prèviament a les claus.

Per escollir aquestes funcions, però, s'han dissenyat dos experiments més. El primer serveix per escollir quina família de funcions de hash es fa servir i el segon per escollir si milloren els resultats si s'hi apliquen altres estratègies de gestió de col·lisions a sobre, concretament rolling hash.

Tots tres experiments s'han dut a terme amb un mateix programa adaptat a les especificitats de cada experiment. Aquest programa té, d'entrada, el número k de funcions, la mida m de la taula, el número n d'iteracions, la mida màxima de les claus i el número de vegades que cal efectuar l'experiment, i dóna la mitjana de falsos positius que han donat les diferents execucions de l'experiment.

Per experimentar amb diferents funcions, es fan servir les famíles de funcions *Mod* ($h_{ab} = (ax + b) \mod p \mod m$) i *Knuth* ($h_{ab} = \frac{(ax+b) \mod 2^w}{(2^{w-M})}$), on p, m, w i M són específics de cada família.

2 Disseny experimental

2.1 Experiment 1

Es calculen els falsos positius d'un filtre de Bloom fent servir 1024 claus aleatòries de 20 dígits alfanumèrics i hashejant-les amb dues famílies de funcions diferents, *Mod* i *Knuth*, que es comparen entre elles. El número de funcions a utilitzar pel filtre es calcula segons la mida m, que és el que es va variant per veure el seu efecte sobre els falsos positius.

2.2 Experiment 2

Els calculen els falsos positius de la mateixa manera que en l'experiment 1, però aquesta vegada es compara la funció que hagi donat menys falsos positius amb fer-la servir per fer rolling hash.

2.3 Experiment 3

Els calculen els falsos positius de la mateixa manera que als experiments 2 i 3, però aquesta vegada es compara la funció que hagi donat menys falsos positius a l'experiment 3 amb aplicar abans a les claus SHA256.

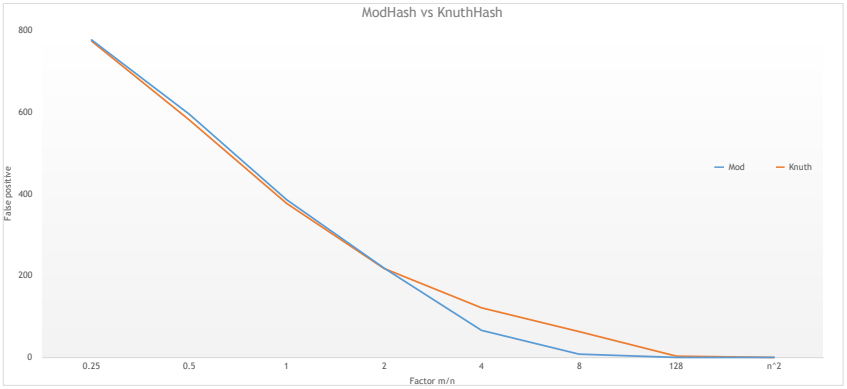

Figura 1: Falsos positius de *Mod* i *Knuth* variant el quocient m/n

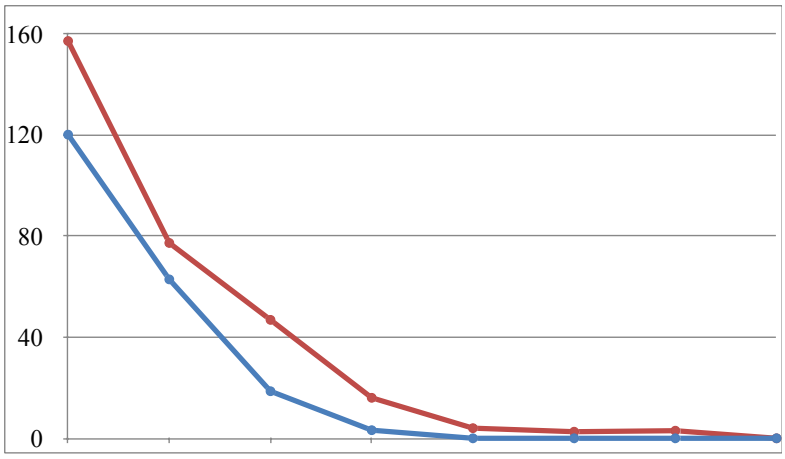

Figura 2: Falsos positius de *Mod*(blau) i *Rolling hash* (vermell) variant el quocient m/n

3 Resultats i conclusions

Es veu a la figura 1 que *Mod*, tot i que no excessivament, supera a *Knuth* en efectivitat. És per això que l'escollim pels següents experiments.

Es veu a la figura 2 que *Mod* funciona millor sense *rolling hash* que amb.

La figura 3 indica que passar-li SHA256 a les claus, com les reparteix uniformement per l'univers de claus, millora notablement el funcionament del filtre de Bloom, sobretot quan hi ha més falsos positius en general, és a dir, quan m/n és més petit.

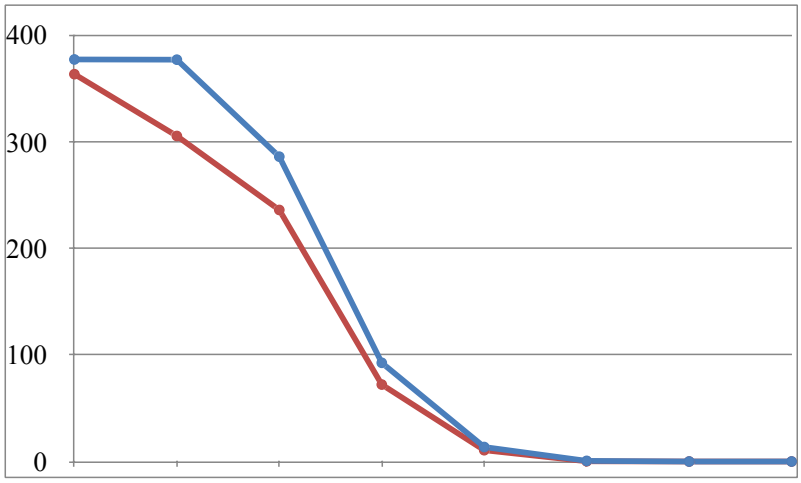

Figura 3: Falsos positius de *Mod* aplicant-li SHA256 (blau) o no (vermell) a les claus prèviament variant el quocient m/n