COMP90056 Assignment A

Tingsheng (Tinson) Lai (731319)

2019

1 Part 1

As we are using a single hash function from a 2-universal hash family and the bitmap width is r, for a single item in the stream, the probability of a cell in the bitmap remains false is

$$1 - \frac{1}{r}$$

By the property of independence, after the m-th item in the stream (i.e., the end of the stream), it becomes

$$\left(1 - \frac{1}{r}\right)^m$$

If an item is reported to be in the stream but it was not actually in the stream, it means such cell was changed to false beforehand, thus the probability of false positive is

$$1 - \left(1 - \frac{1}{r}\right)^m \leq 1 - e^{-\frac{m}{r}}$$

By introducing the parameter $\epsilon > 0$, we let

$$1 - e^{-\frac{m}{r}} \leq \epsilon$$

Solving the inequality, r needs to satisfy that

$$\begin{aligned} 1 - e^{-\frac{m}{r}} &\leq \epsilon \\ e^{-\frac{m}{r}} &\geq 1 - \epsilon \\ -\frac{m}{r} &\geq \log(1 - \epsilon) \\ r &\geq \frac{m}{-\log(1 - \epsilon)} \end{aligned}$$

to achieve the goal.

2 Part 2