Problem 1: Given a universe, U, a threat set, T, and a safe set, S, answer membership queries for $x \in S \cup T$ correctly and for $x \in U \setminus (S \cup T)$ correctly with at most $(1/2)^{m_t - 1}$ chance of error.

Let $s = |S|$, $t = |T|$, m_t the # hashes a Bloom filter on T uses and m_s the # hashes a Bloom filter on S uses.

Assume $s^2 \leq t$.

Construct Bloom filters for both sets. Use $4 m_s t$ space for the Bloom filter on S.

The probability of a wrong answer $x \notin T$ is at most $(1/2)^{m_t}$. The probability of a wrong answer $x \notin S$ is at most $(s/2t)^{m_s}$.

Construct two Bloom filters in $\mathcal{O}(s+t)$ expected time so you can answer queries correctly on $S \cup T$.

First Step? Break down the problem.

1.) Highlight info. into Categories.
2.) Bucket the info.
3.) Sort the buckets.
4.) Restate the problem.

Do this just as we did in the previous problem. (I've already done it for us above).

What similar problems have we seen before?

Hashing without collisions.

How did we solve it?

Keep drawing a hash function until we find a collision free one.

What guarantees do Bloom filters give us?

(Recall the relevant definitions/facts)

$x \in X$ answered correctly.

$x \notin X$ answered correctly with high probability

Idea: Try $x \in S$, if $x \notin S$ then try $x \in T$. (Try simple algorithms, fix them later!)

What guarantees do we need to make this work?

If $x \in T$, then $x \notin S$ is always answered correctly.

This looks exactly like hashing w/o Collisions!

What do we need to ensure we can find the right hash(es) in linear time?

A constant probability of success.

What's the probability of an error?

???

(Easier/Similar Question) What's the expected number of errors?

$$t\,P(\text{error}/\text{collision}) = t\left(s/2t\right)^{m_s} \leq t/\left(2\sqrt{t}\right)^{m_s}$$

Can we bound the probability of any errors by their expectation? If $X \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}[X] &= P(X=0)0 + P(X=1)1 + P(X=2)2 + \dots \\
&= P(X=1) + P(X=2)2 + \dots \\
&\geq P(X=1) + P(X=2) + P(X=3) + \dots
\end{aligned}$$

$$\Longrightarrow \mathbb{E}[X] \geq P(X \geq 1)$$

Let $m_s = 2$, then $P(X \geq 1) \leq \mathbb{E}[X] \leq 1/4$.

Algorithm:

Create Bloom filter for S such that if $x \in T$ then $x \notin S$ is always answered correctly.

Each Bloom filter on S fails w/ at most $1/4$ probability. Takes $O(s)$ ~~ex~~ time to build and $O(t)$ time to check for errors. We build a constant expected number of such Bloom filters, for $O(t+s)$ expected time.

Takes $O(t)$ time to build a Bloom filter for T.

To test membership, check $x \in S$. If $x \notin S$, check $x \in T$.

If $x \in U \setminus (S \cup T)$, then we'll report $x \in T$ w/ at most $1/2^{m_t}$ probability (if the Bloom filter for T fails).