Question 1

(1) Let a_k denote the element of S inserted during k^{th} iteration in the above algorithm. Let ε_k be the event that a_k is inserted in table T_2. Calculate $P(\varepsilon_k)$.
(2) Let X_k denote the random variable defined as follows. If a_k is stored in T_1, then $X_k = 0$; otherwise, it is the number of elements among $\{a_1, ..., a_{k-1}\}$ that collide with it in T_2. Provide a suitable upper bound on $E[X_k|E_k]$.
(3) Use (2) to show that we can design a perfect hashing (at most one element in any location of T_1 and T_2) using n which is asymptotically much smaller than s^2.

(1)

Probability of event ε_k would be the same as the probability of the event that a_k collides with at least one of $a_1, a_2, ..., a_{k-1}$ under hash function h_1. Call this event to be E.
Call the event that a_k collides with a_i to be E_i where $1 \leq i \leq (k-1)$ Now,

$$P(E) = P(\bigcup_{i=1}^{k-1} E_i)$$

Since h_1 is selected randomly uniformly from universal hash family H we get for any $a, b \in U$,

$$P(h_1(a) = h_1(b)) \leq \frac{c}{n}$$

Thus,

$$P(E_i) \leq \frac{c}{n}$$

By union theorem,

$$P(E) \leq P(\sum_{i=1}^{k-1} E_i)$$

$$P(E) \leq \frac{(k-1)c}{n}$$

also we have $P(\varepsilon_k) = P(E)$. Hence we get,

$$P(\varepsilon_k) \leq \frac{(k-1)c}{n}$$

(3)

Question 2

Consider a collection $X_1, ..., X_n$ of n independent geometrically distributed random variables with expected value 2. Let $X = \sum_{i=1}^{N} X_i$ and $\delta > 0$
(1) Derive a bound on $P(X \geq (1+\delta)(2n))$ by appyling the Chernoff bound to a sequence of $(1+\delta)(2n)$ fair coin tosses.
(2) Directly derive a Chernoff like bound on $P(X \geq (1+\delta)(2n))$ from scratch.
(3) Which bound is better?

Let Y be a geometrically distributed random variable then $P(Y = k) = (1-p)^{k-1}p$ and $E[Y] = \frac{1}{p}$.

(1)

Since $E[X_i] = 2$ we find that each X_i is geometrically distributed with parameter $p = \frac{1}{2}$.
As $p = \frac{1}{2}$ we can think of each X_i as the number of coin tosses required to get the first head.
So their sum $X = \sum_{i=1}^{n} X_i$ can be thought of as the number of coin tosses required to get n heads.
We have to find the probability that no less than $(1+\delta)(2n)$ coin tosses are required to get n heads, this is same as the probability that less than n heads appear in $(1+\delta)(2n)$ coin tosses which in turn is same as the probability that no less than $n(1+2\delta)$ tails appear in $(1+\delta)(2n)$ coin tosses
Let Y_i be the random variable taking value 1 when i^{th} coin toss results in a tail and 0 otherwise. Each Y_i is a Bernoulli random variable with probability equal to $\frac{1}{2}$
Consider $Y = \sum_{i=1}^{(1+\delta)(2n)} Y_i$. By linearity of expectation, $E[Y] = (1+\delta)n$
Now,

$$P(X \geq (1+\delta)(2n)) = P(Y \geq n(1+2\delta))$$

$$P(Y \geq n(1+2\delta)) = P\left(Y \geq \frac{(1+\delta)n(1+2\delta)}{1+\delta}\right)$$

$$P(Y \geq n(1+2\delta)) = P\left(Y \geq (1+\delta)n\left(1+\frac{\delta}{1+\delta}\right)\right)$$

$$P(Y \geq n(1+2\delta)) = P\left(Y \geq E[Y]\left(1+\frac{\delta}{1+\delta}\right)\right)$$

Applying chernoff bound on Y,

$$P(Y \geq n(1+2\delta)) \leq \frac{e^{n\delta}(1+\delta)^{n(1+2\delta)}}{(1+2\delta)^{n(1+2\delta)}}$$

$$P(Y \geq n(1+2\delta)) \leq \left(\frac{e^{\delta}(1+\delta)^{(1+2\delta)}}{(1+2\delta)^{(1+2\delta)}}\right)^n$$

Thus we get,

$$P(X \geq (1+\delta)(2n)) \leq \left(\frac{e^{\delta}(1+\delta)^{(1+2\delta)}}{(1+2\delta)^{(1+2\delta)}}\right)^n$$

(2)

$$P(X \geq (1+\delta)(2n)) = P(e^{tX} \geq e^{t(1+\delta)(2n)})$$

By Markov's inequality,

$$P(e^{tX} \geq e^{t(1+\delta)(2n)}) \leq \frac{E[e^{tX}]}{e^{t(1+\delta)(2n)}}$$

Now,

$$E[e^{tX}] = E[e^{(tX_1+...+tX_n)}]$$

$$E[e^{tX}] = E[e^{tX_1}e^{tX_2}...e^{tX_n}]$$

$$E[e^{tX}] = E[\Pi_{i=1}^{n} e^{tX_i}]$$

We know that X_is are independent so e^{tX_i}s are also independent amd hence we get,

$$E[e^{tX}] = \Pi_{i=1}^{n} E[e^{tX_i}]$$

and

$$E[e^{tX_i}] = \sum_{j=1}^{\infty} e^{tj} P(X_i = j)$$

$$E[e^{tX_i}] = \sum_{j=1}^{\infty} e^{tj} (1-p)^{j-1} p$$

$$E[e^{tX_i}] = e^t p \sum_{j=1}^{\infty} (e^t(1-p))^{j-1}$$

We had $p = \frac{1}{2}$. Take $t < \ln 2$ so that $e^t(1-p) < 1$.
By using the formula of summation of a GP,

$$E[e^{tX_i}] = e^t p \frac{1}{1 - e^t(1-p)}$$

$$E[e^{tX}] = E[\Pi_{i=1}^{n} e^{tX_i}]$$

$$E[e^{tX}] = \left(\frac{e^t p}{1 - e^t(1-p)}\right)^n$$

Thus we get,

$$P(e^{tX} \geq e^{t(1+\delta)(2n)}) \leq \frac{1}{e^{t(1+\delta)(2n)}} \left(\frac{e^t p}{1 - e^t(1-p)}\right)^n$$

$$P(e^{tX} \geq e^{t(1+\delta)(2n)}) \leq \left(\frac{e^t p}{(1 - e^t(1-p))(e^{2t(1+\delta)})}\right)^n$$

This is true for every value of $t < \ln 2$ and hence differentiating it to find the minimum value -

$$\frac{d}{dt}\left(\frac{e^t p}{(1 - e^t(1-p))(e^{2t(1+\delta)})}\right)^n = 0$$

$$\frac{d}{dt}\frac{e^t p}{(1 - e^t(1-p))(e^{2t(1+\delta)})} = 0$$

$$-(1 - e^t(1-p))(1+2\delta)(e^{-t(1+2\delta)}) + (e^{-t(1+2\delta)})(e^t(1-p)) = 0$$

$$e^t(1-p) = (1+2\delta)(1 - e^t(1-p))$$

$$e^t(1-p) = \frac{1+2\delta}{2+2\delta}$$

Put $p = \frac{1}{2}$,

$$e^t = \frac{1+2\delta}{1+\delta}$$

So we get after substituting this value of e^t,

$$P(e^{tX} \geq e^{t(1+\delta)(2n)}) \leq \left(\frac{(1+\delta)^{2+2\delta}}{(1+2\delta)^{1+2\delta}}\right)^n$$

So,

$$P(X \geq (1+\delta)(2n)) \leq \left(\frac{(1+\delta)^{2+2\delta}}{(1+2\delta)^{1+2\delta}}\right)^n$$

(3)

Let the first and second bounds are $B1$ and $B2$ respectively,

$$\frac{B1}{B2} = \frac{e^\delta}{1+\delta}$$

$$\frac{B1}{B2} \geq 1$$

Thus, the second bound is better.