CSCI 7000: Advanced Data Structures (Assignment 1)

Nicholas Papadopoulos

September 2, 2022

Problem 1

Part A

$$x_1 \oplus x_2 = x_1 \oplus x_3$$
$$x_1 \oplus x_1 \oplus x_2 = x_1 \oplus x_1 \oplus x_3$$
$$x_2 = x_3$$

So, we need to find the probability that $x_2 = x_3$. There are 2^l possibile bitsrings of length l, so the probability that the same bitstring is randomly generated for both x_2 and x_3 is $\frac{\mathbf{1}}{\mathbf{2^l}}$.

Part B

Set $x_1 \oplus x_2$ can map to some other random string, say s. Each bitstring, s, of length l can have 2^l pairs, where order matters, that xor to s. This is because you can choose any random bitstring, r, of length l, xor it with s, and the result, r' will be the paired value of r where $roplusr' = s$. In other words, any random string out of the possible 2^l strings can serve as x_3 with a deterministic, corresponding x_4.

Hence, x_1, x_2, and x_3 can be anything without consequence. With these set, however, we know that $x_4 = x_1 \oplus x_2 \oplus x_3$. In other words, there is only one possible solution for x_4 once the free variables x_1, x_2, and x_3 are chosen. Since x_4 is chosen randomly, and there are 2^l possible choices, the probability that $x_1 \oplus x_2 = x_3 \oplus x_4$ is $\frac{\mathbf{1}}{\mathbf{2^l}}$.

Part C

First, we can determine the probability that x maps to the same keys for A and B. That is, what is the probability that $x_1 = x_2$? We have determined that this probability is $\frac{1}{2^l}$ in part (A).

$$h_{A,B}(x) = A[x_1] \oplus B[x_1]$$
$$h_{A,B}(y) = A[y_1] \oplus B[y_1]$$

We can now ask what the probability is that $A[x_1] \oplus B[x_1] = A[y_1] \oplus B[y_1]$. Since $A[x_1]$, $A[x_2]$, $A[y_1]$, and $A[y_1]$ are random l-bit strings, we can use our answer in part (B) to say that the probability of collision is $\frac{1}{2^l}$.

Problem 2

$$\begin{aligned}
h_1(x) &= (x \mod 6) \mod 4 \\
h_2(x) &= (2x \mod 6) \mod 4 \\
h_3(x) &= (3x \mod 6) \mod 4 \\
h_4(x) &= (4x \mod 6) \mod 4 \\
h_5(x) &= (5x \mod 6) \mod 4
\end{aligned}$$

Here, we can see that we can pick h_3 and see that

$$h_3(x) = \begin{cases} 0 & x \text{ is even} \\ 3 & x \text{ is odd} \end{cases}$$

This is because 3 times any even number will be exactly divisible by 6, giving a remainder of 0, and 3 times and odd number will be 3 times and even number plus 3, giving a remainder of 3. Since both 0 and 3 are less than 4, modding by 4 does not effect the result. Hence, $\frac{1}{2}$ off the possible keys to h_3 will result in the same answer, so there the probability of a collision is $\mathbf{\frac{1}{2}} > \mathbf{\frac{1}{4}}$.

Problem 3

We can take the binary representation of each character and mod them by some number, say 10. Then we can multiply each character by subsequent powers of ten so that no character can interfere with the next when added together, since the highest value of some power, p, $9 * 10^p < 10^{p+1}$. This holds true for any base, say a, not only 10. We can use this fact to include a salt, so we can say that

$$a = b + \texttt{salt}$$

Where b is some minimum base number we would like to utilize. We can then define a character encoding by

$$x_n = c_n \mod a$$

We then have that

$$h(s) = x_1(a^n) + x_2(a^{n-1}) + ...x_n(a^0)$$

$h(t)$ would then be

$$h(t) = x_2(a^n) + x_3(a^{n-1}) + ...x_{n+1}(a^0)$$

So, to retrieve $h(t)$ given s, t, and $h(s)$, we simply subtract $x_1(a^n)$, multiply the result by a, then add x_{n+1}.

$$h(t) = (h(s) - x_1(a^n))a + x_{n+1}$$

This uses constant time operations for all portions of this equation, consisting of modulo, exponentiation, subtraction, multiplication, and addition.