CS191 – Fall 2014
Homework 5: due in lecture Oct. 22nd

1. ***The six state protocol.*** An alternative to the BB84 protocol is the six state protocol in which Alice and Bob have three bases to choose from when encoding and measuring. That is, Alice uniformly chooses a basis out of the σ_z, σ_x and σ_y bases and then sends one of the two orthogonal states in the chosen basis. Explicitly, Alice sends either $\{|0\rangle \text{ or } |1\rangle\}$ or $\{|+\rangle \text{ or } |-\rangle\}$ or $\{|+i\rangle \text{ or } |-i\rangle\}$, where $|\pm i\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i\,|1\rangle)$ are the eigenstates of σ_y. Similarly Bob has three choices from which to uniformly choose a basis to measure in: σ_z, σ_x or σ_y.

   Assuming that Eve performs an intercept resend attack on every qubit (*i.e.*, she measures in a uniformly chosen random basis out of the three and sends her resulting state onto Bob), what is the error rate Alice and Bob will see during the error estimation stage.

2. ***Entanglement-based protocol.*** Entanglement-based protocols begin with a trusted source (which could be Alice herself) distributing one qubit of a two-qubit entangled state to Alice and Bob. Then Alice and Bob do measurements in some choice of basis on the qubit they receive. The sifting stage consists of keeping the outcomes of the rounds where there is agreement between their basis choices.

   Let the entangled state that is distributed be

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B),$$

   where the subscript denotes whether that qubit is sent to Alice or Bob. Suppose Alice and Bob both measure in σ_z or σ_x basis with equal probability. Further, they agree that getting measurement result $|0\rangle$ / $|1\rangle$ when measuring in the σ_z basis will be denoted as 0 / 1 bit value. Similarly, they agree that getting measurement result $|+\rangle$ / $|-\rangle$ when measuring in the σ_x basis will be denoted as 0 / 1 bit value.

   (a) Write the state $|\Psi\rangle$ in the basis $\{|+\rangle_A \otimes |+\rangle_B, |-\rangle_A \otimes |+\rangle_B, |+\rangle_A \otimes |-\rangle_B, |-\rangle_A \otimes |-\rangle_B\}$ so that it is easy to compute the outcomes and their respective probabilities when Alice and/or Bob measure in the σ_x basis.

   (b) Complete the following table, which is the similar to the one we did in class, except this time for this entanglement-based protocol. The first column is done for you. If any entry is random then choose a value at random but make sure the other entries in that column are consistent with the choice you make.

TABLE I: default

Alice basis choice	σ_z	σ_x	σ_x	σ_z	σ_z	σ_z	σ_x	σ_z	σ_x	σ_z
Bob basis choice	σ_x	σ_z	σ_x	σ_x	σ_z	σ_x	σ_z	σ_z	σ_x	σ_z
Alice measurement result	$\|0\rangle$									
Alice bit value	0									
Bob measurement result	$\|+\rangle$									
Bob bit value	0									
Keep after sifting?	No									

   (c) Now suppose that as a result of noise or Eve's intervention, the state that actually gets to Alice and Bob is

$$\rho = \frac{p}{4} I_4 + (1-p)\,|\Psi\rangle\langle\Psi|\,,$$

   where I is the 4×4 identity matrix.

   i. Suppose Alice measures in the σ_x basis and gets result $|+\rangle$. What are the probabilities for Bob's possible outcomes if he also measures in the σ_x basis.

   ii. Suppose Alice measures in the σ_x basis and gets result $|+\rangle$. What are the probabilities for Bob's possible outcomes if he measures in the σ_z basis.

3. ***Universal hashing for privacy amplification.*** We discussed in class how privacy amplification can be performed by applying a function from a universal hashing family to Alice's and Bob's reconciled bit string.

A hash function, h, is a map from a set U (which could be infinite in cardinality) to a finite set $\{1, ...M\}$. The domain and range of the hash functions we are interested in are finite bit strings, *i.e.*, $h : \{0,1\}^n \to \{0,1\}^m$ with $n > m$. A family of hash functions is a collection of such functions parameterized in some fashion.

Definition. $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^m, n > m\}$ is a universal family of hashing functions if for all $x_1 \neq x_2 \in \{0,1\}^n$ and $y \in \{0,1\}^m$,

$$\Pr_{h\in\mathcal{H}}\{h(x_1) = h(x_2) = y\} \leq \frac{1}{2^m},$$

when h is sampled uniformly from $\mathcal{H}$.

This definition says that the probability of a *collision* when we map an n-bit string to an m-bit string using a hash function from this family is at most $\frac{1}{2^m}$. It is pretty amazing that such function families exist even when n is much larger than m.

A useful universal hashing family is defined through multiplication by the set of $m \times n$ matrices with random binary entries (the matrix multiplication is done modulo 2, or more technically, over the finite field $GF(2)$). That is, $h_M(x) = Mx$, where M is randomly chosen from $\mathcal{M}_{m\times n}(2)$, the set of $m \times n$ matrices over $GF(2)$. However, for QKD, there is a more efficiently specified class of universal hashing functions that is more commonly used, and these are defined through multiplication by Toeplitz matrices over $GF(2)$. That is, $h_T(x) = Tx$, with T being a $m \times n$ Toeplitz matrix specified by $n+m-1$ uniform random bits (see below for a definition of Topelitz matrices).

So after the reconciliation stage, when Alice and Bob share an n-bit string, $\mathbf{x}$, and have estimated that Eve has k bits of knowledge about that string, Bob randomly choose a $m \times n$ Topelitz matrix, T, (where $m \leq n-k$) and communicates it to Alice. Then they both calculate the privacy amplified string of length m by computing $\mathbf{y} = T\mathbf{x}$. This is secure, even if Eve has full knowledge of T. It is even secure in practice if Eve has greater than k bits of knowledge, although the mathematical guarantee is only for when she has k bits or less.

We will work through an example in this problem. Let $n = 10$, and $\mathbf{x} = (1,0,0,1,1,1,0,1,0,0)^\mathsf{T}$. Then suppose $k = 3$, and Bob chooses the following 7×10 Toeplitz matrix to define the hash function, and communicates it to Alice:

$$\begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{pmatrix}$$

(a) Compute $\mathbf{y} = T\mathbf{x}$, Alice and Bob's privacy amplified bit string.

(b) Flip $l = 4$ random bits in $\mathbf{x}$ and write out this $\mathbf{x}'$. Let this represent Eve's best guess at $\mathbf{x}$, and compute $\mathbf{y}' = T\mathbf{x}'$. What is the hamming distance between $\mathbf{y}$ and $\mathbf{y}'$ (see below for a definition of hamming distance)?

(c) Repeat (b) for $l = 3, 2, 1$. Did you find any instance where $\mathbf{y}' = \mathbf{y}$? Are you convinced that even if Eve knows some set of bits and T, that she has little information about the privacy amplified string $\mathbf{y}$?

Some definitions

1. A *Toeplitz matrix* is one of the form:

$$\begin{pmatrix} a & b & c & d \\ e & a & b & c \\ f & e & a & b \\ g & f & e & a \end{pmatrix}$$

   That is, each row is a right-shifted version of the previous one, with a new element in the first slot. We see that while a general $n \times m$ matrix is specified by nm entries, a Toeplitz matrix is specified by $n+m-1$ entries (basically the first row and column). This is a useful property for QKD since it means that the choice of hash function (which is equivalent to specifying a Topelitz matrix) can be communicated between Alice and Bob by specifying $n+m-1$ bits.

2. The *hamming distance* between two bit strings $\mathbf{x_1}$ and $\mathbf{x_2}$, $d_H(\mathbf{x_1}, \mathbf{x_2})$, is the number of places where they differ in value. For example,

$$d_H((1,0,1,1,0,0),(0,0,1,0,1,0)) = 3$$