CSC 591 Cryptography 11/12/2018

Lecture 16: Zero-Knowledge Part I

Lecturer: Alessandra Scafuro *Scribe: Christian Morris*

Topic/Problem

In previous lectures, we have only concerned ourselves with situations where an honest Alice communicates with an honest Bob in the presence of an eavesdropper. Consider the situation where neither Alice nor Bob trusts the other communicating party - how is it possible that Alice (who we'll refer to as the prover) can convince Bob (considered henceforth as the verifier) to trust her?
In this lecture we discuss examples of zero-knowledge proofs, and later define what it means for something to be a zero-knowledge proof.

Definition By Example: Where's Waldo?

Suppose there is a puzzle from "Where's Waldo?" in which Alice knows the location of Waldo, but Bob does not. Alice would like to convince Bob that she knows the location of Waldo, but would not like to reveal any information to Bob that would help him locate Waldo on his own - Alice would only like to convince Bob that she knows where Waldo is. How can this be done?

Solution: Alice places a large, opaque piece of cardboard over the entire picture, with a hole in the center that reveals Waldo.

With this solution, Bob is not able to gain any information for himself about where Waldo is within the puzzle, but Alice has proven to Bob that she does in fact know the solution herself.
This allows us to have a loose definition of what a zero-knowledge proof is:

Loose Definition:
A zero-knowledge proof for a language $\mathcal{L}$ is an interactivve protocol executed between a prover P and a verifier V, where the prover wants to convince the verifier that a fact (*a theorem*) is true.

Interactive Proof By Example: A Lady Testing Tea

Suppose Muriel tells Ronald that tea when poured into milk tastes different than milk poured into tea, but Ronald would like proof that Muriel can tell the difference in taste. An interactive proof can be built as follows:

1. In private, Ronald flips a coin to decide whether tea or milk is poured first, and then gives the cup to Muriel

2. Muriel is left to guess which was poured first
    - If Muriel can really tell whether milk or tea was poured first, she gets it right.
    - If there is no difference in the two kinds of tea, Muriel guesses correctly with probability $\frac{1}{2}$.
3. Repeat this testing phase n times.
    - If there's no difference in the two kinds of teas, Muriel will guess correctly for all cups of tea with probability $\frac{1}{2^n}$.

With the above example, we are able to craft a more precise definition of what it means for something to be zero-knowledge:

Clever Definition:
An interaction is zero-knowledge if Bob is able to generate transcripts without interacting with Alice.

This outlines the fact that Bob is convinced by Alice through *how* the transcript was generated (in response to his challenges) as opposed to *what* the transcript was.

Formal Definition

Given the two examples above, we are able to construct a formal definition of a zero-knowledge proof.

Definition [GMR 1985]:
A zero-knowledge proof is an interactive protocol satisfying:

- The prover can always convince the verifier of any true statement
- The verifier can't be convinced of a false statement (even by a cheating prover), except with very low probability
- There is an efficient procedure to output "same-looking" protocol transcripts

Sudoku Zero-Knowledge Proof

Alice wants to convince Bob that there is a solution to a Sudoku puzzle, but doesn't want to reveal the solution to Bob so that he can solve it on his own. The zero-knowledge protocol is the following:

Zero-knowledge protocol:

1. Alice randomly relabels {1,...,9}
2. Alice writes relabeled solution on a scratch card, and shows the card to Bob
3. Bob asks Alice to scratch off one of the following:

- A particular row
- A particular column
- A particular 3 x 3 block
- Initial positions

and checks the consistency

4. Repeat process n times

Sudoku Zero-Knowledge Proof: Analysis

Observe that if Alice is able to successfully answer all challenges, her scratch card satisfies:

- Every row, column, and block is a permutation of {1,...,9}
- Initial positions are consistent with relabeling of {1,...,9} in the original puzzle

Then the puzzle has a solution.

It is acknowledged that in a single iteration of this protocol that if Alice is cheating she has a probability $\leq \frac{27}{28}$ of being undetected. This protocol is still appropriate given that the protocol is repeated n times. For example, if $n = 2500$, Alice would be caught cheating with 99% probability.

Looking further into how the transcript for this protocol is formed, so long as Alice is following the protocol and there is a solution, each round's transcript is one of the following:

- A random permutation of {1,...,9} in a random row,
- A random permutation of {1,...,9} in a random column,
- A random permutation of {1,...,9} in a random block, or
- A random relabeling of the original puzzle's initial positions

All of these transcripts can be generated by Bob himself without the solution.

Zero-Knowledge Proofs for Everything?

In finding a zero-knowledge proof protocol for Sudoku, a theorem was developed:

Theorem [Yato 2003]:
$n \times n$ **Sudoku is NP-complete.**

In the development of this theorem, a more generalized theorem emerged:

Theorem:
Every **NP statement can be proven in zero-knowledge.**