CS 127/CSCI E-127: Introduction to Cryptography

Problem Set 10

Assigned: Nov. 27, 2013 Due: Dec. 6, 2013 (5:00 PM)

Justify all of your answers. See the syllabus for collaboration and lateness policies. Submit solutions by email to `mbun@seas` (and please put the string "CS127PS10" somewhere in your subject line).

Problem 1. (Public-Key Identification Schemes) On Problem Set 6, you showed how to construct a protocol that allows a user to identify herself to a server, provided the user and the server share a secret key k. Here you will see how zero-knowledge proofs can be used to construct *public-key* identification schemes, where the user has a secret key sk and publishes a public key pk. The server only needs to know the public key of the user to verify the user's identity. Similarly to PS6, for security we require that an adversary who is given the public key pk and engages in polynomially many executions of the identification protocol with the user still cannot successfully impersonate the user, except with negligible probability.

Here is a construction of such a public-key identification scheme based on the zero-knowledge proof for Quadratic Residuosity given in class. On security parameter n, the user generates her keys by picking two random n-bit primes p_1, p_2, computing $N = p_1 \cdot p_2$, choosing $q \xleftarrow{R} \mathbb{Z}_N^*$, computing $x = [q^2 \bmod N]$, and setting $pk = (N, x)$ and $sk = (N, q)$.

A single execution of the identification protocol between the user U and a server S proceeds as follows.

1. $U = U(1^n, sk)$ and $S = S(1^n, pk)$ run n sequential executions of the zero-knowledge proof for Quadratic Residuosity on input (N, x), with S playing the role of the verifier and U playing the role of the prover with **NP** witness q.

2. S accepts if all executions of the zero-knowledge proof are accepting.

You will justify the security of this identification scheme in 3 steps.

a) Argue that the zero-knowledge proof for Quadratic Residuosity on input (N, x) not only proves that $x \in \mathrm{QR}_N$, but actually that the prover "knows" a square root of x. To make this more precise, suppose P^* is a prover strategy that convinces the verifier to accept with probability at least $1/2 + \varepsilon$ in one execution of the Quadratic Residuosity interactive proof, for a constant $\varepsilon \in (0, 1/2)$. Show how to use P^* to obtain a square root of x modulo N in polynomial time. (Hint: first argue that P^* can answer both challenges at least an ε fraction of the time.)

b) Suppose that an adversary $A = A(1^n, pk)$ (playing the role of U) convinces the server S to accept in a single execution of the identification scheme above with nonnegligible probability (over the choice of pk and the randomness in the interactive proof). Using the previous item, explain how A can be used to invert Rabin's function (and thus factor) with nonnegligible probability. (You do not need to give a full formal proof here, since the details are rather messy.)

c) Now use the zero-knowledge property of the QUADRATIC RESIDUOSITY interactive proof to argue that engaging in polynomially many executions of the identification scheme with the honest user $U = U(1^n, sk)$ does not enable an adversary $A = A(1^n, pk)$ to later impersonate U, except with negligible probability (under the Factoring Assumption).

Remark: The "Fiat-Shamir heuristic" is a method for converting identification schemes such as the above into digital signature schemes. The public key and secret key of the signature scheme are as in the identification scheme. To sign a message m, the $U = U(1^n, sk)$ runs an execution of the ID scheme on its own, but uses a hash of the message and the transcript t of previous messages in the ID scheme, $H(m\|t) \in \{0,1\}^n$, for the random challenge bits of the verifier in each of the n executions of the QUADRATIC RESIDUOSITY interactive proof. The intuition is that if H is a strong enough hash function, then an adversary will have no control over these challenge bits (even given hash values on other messages), so it is just like running the actual interactive proof. The resulting signature scheme is conjectured to be secure (and this intuition is supported by a proof of security in the idealized random oracle model). This heuristic yields some of the most efficient digital signature schemes used in practice.

Problem 2. (Private Information Retrieval) Consider a server which holds a database D consisting of m bits of data. The goal of *private information retrieval* is to allow a user to retrieve an entry of D without revealing to the server which bit was requested. A simple way to solve this problem is to have the server just send the entire database D whenever a user makes a request. But this amount of communication between the server and user can be prohibitive when m is large. In this problem, you will show how to use homomorphic encryption to solve this problem with much less than m bits of communication between the user and the server.

The Goldwasser-Micali encryption scheme (KL1e §11.1) is a public-key encryption scheme for message space $\mathcal{M} = \{0,1\}$ based on the conjectured hardness of distinguishing squares from non-squares modulo a composite.

- $\mathsf{Gen}(1^n)$: Sample two n-bit primes p, q, each of which is congruent to 3 mod 4. The public key is $pk = N = pq$ and the secret key is $sk = (p, q)$.
- $\mathsf{Enc}_{pk}(m)$ for $m \in \{0,1\}$: Sample $x \xleftarrow{R} \mathbb{Z}_N^*$. If $m = 0$, output the ciphertext $c = [x^2 \bmod N]$. If $m = 1$, output the ciphertext $c = [-x^2 \bmod N]$.
- $\mathsf{Dec}_{sk}(c)$: Using p and q, determine whether c is a quadratic residue modulo N. If $c \in \mathrm{QR}_N$, output 0, and otherwise output 1.

For this problem, assume the correctness and security (indistinguishable encryptions) of this encryption scheme.

a) Show that the Goldwasser–Micali encryption scheme is strongly homomorphic with respect to XOR (see Lecture Notes 20 for a definition).

b) For a string $r \in \{0,1\}^k$, define $f_r : \{0,1\}^k \to \{0,1\}$ as the function $f_r(m) = [\sum_i r_i m_i \bmod 2]$. Show that the Goldwasser–Micali encryption scheme is strongly homomorphic with respect to the functions f_r. The homomorphic evaluation function $\mathsf{Eval}_{pk}^{f_r}$ should take $r \in \{0,1\}^k$ as input and run in time $\mathrm{poly}(k, n)$.

Consider the following solution to the private information retrieval problem.

- View the database D as a $\sqrt{m} \times \sqrt{m}$ matrix of bits.
- The user runs $(pk, sk) \xleftarrow{R} \mathsf{Gen}(1^n)$ for the Goldwasser–Micali encryption scheme,.
- If the user wishes to retrieve a bit in the i'th column of D, she constructs ciphertexts $(c_1, \ldots, c_{\sqrt{m}})$ where $c_i \xleftarrow{R} \mathsf{Enc}_{pk}(1)$ and $c_{i'} \xleftarrow{R} \mathsf{Enc}_{pk}(0)$ for all $i' \neq i$.
- The user sends $(pk, c_1, \ldots, c_{\sqrt{m}})$ to the server.
- For each row $r_j \in \{0,1\}^{\sqrt{m}}$ of the database, the server computes a ciphertext $d_j \xleftarrow{R} \mathsf{Eval}_{pk}^{f_{r_j}}(c_1, \ldots, c_{\sqrt{m}})$.
- The server sends $(d_1, \ldots, d_{\sqrt{m}})$ back to the user.

Note that the user and server communicate $O(\sqrt{m} \cdot n)$ bits in this protocol, which can be much smaller than m for a large database.

c) Show that the user can compute any bit in the ith column of D using the ciphertexts received from the server.

d) Formally define what it means for the user to keep her index $(i, j) \in \{1, \ldots, \sqrt{m}\}^2$ private from the server. Prove that the scheme satisfies your security definition.

e) Using a *fully* homomorphic encryption scheme, describe a private information retrieval protocol that communicates only $\mathrm{poly}(n)$ bits (independent of m!). Briefly explain the correctness and security of your protocol (full proof not needed).

f) **Extra Credit:** Using just Goldwasser–Micali encryption, exhibit a private information retrieval protocol with communication $\mathrm{poly}(n) \cdot m^{\varepsilon}$ for an arbitrarily small constant ε. (Hint: use recursion.)