Schnorr Signatures are Non-Malleable in the Random Oracle Model

Andrew Poelstra

12 Feb 2014

Schnorr signatures. The Schnorr signature cryptosystem over a group G, $|G| = q$, is defined as follows. Let $g \in G$ be some generator. Let H be a hash function, modelled as a random oracle, whose image is $\{0,\dots,x-1\}$. All of G,q,g,H are parameters of the cryptosystem and considered public knowledge.

- *Key generation.* Choose $x \in \{1,\dots,q-1\}$ randomly. Then g^x is the public key, x is the secret key.
- *Signing.* Let m be the message to sign. Choose $k \in \{1,\dots,q-1\}$ randomly. Let $e = H(m||g^k)$, $s = k - xe$. Then (e,s) is the signature.
- *Verification.* Given (e,s), compute $g^k = (g^x)^e g^s$. (Note that k is unknown to the verifier, we are just calling this g^k for consistency with the previous step.) Then $H(m||g^k)$ can be calculated and confirmed to be e.

Malleability. We consider the advantange of a *malleating adversary* $\mathscr{A}$ to be the probability that $g^{s'}g^{xe'} = r'$ and $e' = H(m||r')$, where (s',e') is produced by $\mathscr{A}$ given a message m and valid signatures (s_i,e_i), $i = 1,\dots,n$, for m. We require $(s',e') \neq (s_i,e_i)$ and allow $\mathscr{A}$ to choose n.

Theorem 1. *A malleating adversary $\mathscr{A}$ with non-negligible advantage ε can be used to construct an ordinary forging adversary $\mathscr{B}$ with advantage ε.*

Proof. We first demonstrate that if $(s',e') \neq (s_i,e_i)$, then we must have $e' \neq e_i$. To this end, suppose that $H^A(m||r') = e' = e_i = H^A(m||r_i)$. Then since H^A is a random oracle we must have $r' = r_i$ except with negligible probability. But since $g^{s_i} = (g^x)^{e_i} r = (g^x)^{e'} r' = g^{s'}$ we must have $s_i = s'$. This contradicts $(s',e') \neq (s_i,e_i)$. (The point of this comment is that $\mathscr{A}$ is forced to consult the oracle H to compute e'; he cannot simply modify s_i.)

Then $\mathscr{B}$ operates by running $\mathscr{A}$. The hash function that $\mathscr{A}$ sees is a random oracle H^A controlled by $\mathscr{B}$. Suppose we are given a public key g^ℓ and message m, and that $\mathscr{B}$'s goal is to output a valid signature (S,E) such that $g^S(g^\ell)^E = R$ where $H(m||R) = E$. $\mathscr{B}$ operates as follows.

1. First, $\mathscr{A}$ chooses n requests n valid signatures (s_i,e_i) from $\mathscr{B}$. To respond to each query, $\mathscr{B}$ chooses a pair (s_i,e_i) at random from $\{0,\dots,q-1\}^2$. Also, $\mathscr{B}$ sets $H^A(m||g^s(g^\ell)^e) = e$ so

that $\mathscr{A}$ will view this as a valid signature under the public key g^{ℓ}. Notice that since e_i is chosen uniformly at random, this is consistent with $\mathscr{A}$'s view that H^A is a random oracle.

2. Next, $\mathscr{A}$ generates a malleated signature (s', e'). Write $r = g^{s'}(g^{\ell})^{e'}$. If (s', e') does not satisfy $H^A(m||r)$, then $\mathscr{B}$ quits; the attack fails. This occurs with probability $1 - \varepsilon$.

   Otherwise, since $e' \neq e$ and $e' = H^A(m||r)$, to produce e' with non-negligible probability $\mathscr{A}$ must call H^A with input $m||r$. $\mathscr{B}$ responds to this query with $H(m||r)$, that is, $\mathscr{B}$ gives $\mathscr{A}$ the "real" hash of $m||r$.

3. At this point, we claim that the pair (s', e') is a valid forged signature of m. To see that this is so, notice that
$$H(m||g^{s'}(g^{\ell})^{e'}) = H(m||r) = H^A(m||r) = e'.$$

This completes the proof.

□