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 $\{0, \ldots, x - 1\}$. All of $G,q,g,H$ are parameters of the cryptosystem and considered public knowledge.

- **Key generation.** Choose $x \in \{1, \ldots, q - 1\}$ randomly. Then $g^x$ is the public key, $x$ is the secret key.

- **Signing.** Let $m$ be the message to sign. Choose $k \in \{1, \ldots, q - 1\}$ 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 advantage of a *malleating adversary* $\mathcal{A}$ to be the probability that $g^{s'} g^{xe'} = r'$ and $e' = H(m||r')$, where $(s', e')$ is produced by $\mathcal{A}$ given a message $m$ and valid signatures $(s_i, e_i)$, $i = 1, \ldots, n$, for $m$. We require $(s', e') \neq (s_i, e_i)$ and allow $\mathcal{A}$ to choose $n$.

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

**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 $g^{s'} = (g^x)^e r = (g^x)^e r' = g^{s'}$ we must have $s_i = s'$. This contradicts $(s', e') \neq (s_i, e_i)$. (The point of this comment is that $\mathcal{A}$ is forced to consult the oracle $H$ to compute $e'$; he cannot simply modify $s_i$.)

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

1. First, $\mathcal{A}$ chooses $n$ requests $n$ valid signatures $(s_i, e_i)$ from $\mathcal{B}$. To respond to each query, $\mathcal{B}$ chooses a pair $(s_i, e_i)$ at random from $\{0, \ldots, q - 1\}^2$. Also, $\mathcal{B}$ sets $H^A(m||g^e (g^e)^{s'}) = e$ so
that \( A \) will view this as a valid signature under the public key \( g^e \). Notice that since \( e_i \) is chosen uniformly at random, this is consistent with \( A \)'s view that \( H^A \) is a random oracle.

2. Next, \( A \) generates a malleated signature \( (s', e') \). Write \( r = g^e \cdot (g^{e'})^e \). If \( (s', e') \) does not satisfy \( H^A(m||r) \), then \( B \) 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 \( A \) must call \( H^A \) with input \( m||r \). \( B \) responds to this query with \( H(m||r) \), that is, \( B \) gives \( A \) 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

\[
H(m||g^{s'}(g^{e'})^e) = H(m||r) = H^A(m||r) = e'.
\]

This completes the proof. \( \square \)