In many cases, we want a stronger property: the prover actually "knows" why a statement is true (e.g., it knows a "witness")

For instance, consider the following language:

\[ L = \{ h \in G | \exists x \in \mathbb{Z}_p: h = g^x \} = G \]

Note: this definition of \( L \) implicitly defines an NP relation \( R: \)

\[ R(h, x) = 1 \iff h = g^x \in G \]

In this case, all statements in \( G \) are true (i.e., contained in \( L \)), but we can still consider a notion of proving knowledge of the discrete log of an element \( h \in G \) — conceptually stronger property than proof of membership.

Philosophical question: What does it mean to "know" something?

If a prover is able to convince an honest verifier that it "knows" something, then it should be possible to extract that quantity from the prover.

Definition. An interactive proof system \((P, V)\) is a proof of knowledge for an NP relation \( R \) if there exists an efficient extractor \( E \) such that for any \( x \) and any prover \( P^* \)

\[
Pr[w \leftarrow E^*_p(x) : R(x, w) = 1] \geq Pr[(P^*, V)(x) = 1] - \varepsilon
\]

more generally, could be polynomially smaller

knowledge error

Trivial proof of knowledge: prover sends witness in the clear to the verifier

\[ \Rightarrow \text{In most applications, we additionally require zero-knowledge} \]

Note: knowledge is a strictly stronger property than soundness

\[ \Rightarrow \text{if protocol has knowledge error } \varepsilon \Rightarrow \text{it also has soundness error } \varepsilon \text{ (i.e. a dishonest prover convinces an honest verifier of a false statement with probability at most } \varepsilon) \]

Proving knowledge of discrete log (Schnorr's protocol)

Suppose prover wants to prove it knows \( x \) such that \( h = g^x \) (i.e., prover demonstrates knowledge of discrete log of \( h \) base \( g \))

\[
\begin{array}{c}
\text{prover} \\
\hline
r \in \mathbb{Z}_p \\
u \leftarrow g^r \\
z = r + cx \\
\end{array}
\quad
\begin{array}{c}
\text{verifier} \\
u \leftrightarrow c \\
z \leftarrow c \cdot s \mod p \\
\end{array}
\quad
\begin{array}{c}
\text{verify that } g^z = u \cdot h^c \\
\end{array}
\]

assume \( g, h \in G \) where \( G \) has prime order \( q \)
Completeness: if \( z = r + cx \), then 
\[
\begin{align*}
    g^z &= g^{r + cx} \\
    &= g^r g^{cx} = u \cdot h^c
\end{align*}
\]
zero knowledge only required to hold against an honest verifier (e.g., view of the honest verifier can be simulated)

\[ \begin{array}{c}
\text{Honest-Verifier Zero-Knowledge: build a simulator as follows (familiar strategy: run the protocol in reverse):} \\
\text{on input } (g, h); \\
1. \text{Sample } z \leftarrow Z_p \\
2. \text{Sample } c \leftarrow Z_p \text{ and output } (u, c, z) \\
3. \text{Set } u = \frac{g^z}{h^c} \\
\end{array} \]
\[
\left\{ \begin{array}{l}
\text{uniformly random challenge} \\
\text{chosen so that} \\
\text{uniformly random} \\
\text{relates satisfied by } a) \\
\text{valid proof} \\
\end{array} \right. \\
\text{simulated transcript is identically distributed} \\
\text{as the real transcript with an honest verifier} \\
\]

What goes wrong if the challenge is not sampled uniformly at random (i.e., if the verifier is dishonest)

Above simulation no longer works (since we cannot sample \( z \) first)

\( \rightarrow \) To get general zero knowledge, we require that the verifier first commit to its challenge (using a statistically hiding commitment)

\[ \begin{array}{c}
\text{Knowledge: Suppose } P^* \text{ is (possibly malicious) prover that convinces honest verifier with probability } 1. \text{ We construct an extractor as follows:} \\
\text{1. Run the prover } P^* \text{ to obtain an initial message } u. \\
\text{2. Send a challenge } c_1 \leftarrow Z_p \text{ to } P^*. \text{ The prover replies with a response } z_1. \\
\text{3. "Rewind" the prover } P^* \text{ so its internal state is the same as it was at the end of Step 1. Then, send another } \\
\text{challenge } c_2 \leftarrow Z_p \text{ to } P^*. \text{ Let } z_2 \text{ be the response of } P^*. \\
\text{4. Compute and output } \chi = (z_1 - z_2)(c_1 - c_2) \in Z_p. \\
\end{array} \]

Since \( P^* \) succeeds with probability 1 and the extractor perfectly simulates the honest verifier's behavior, with probability 1, both \((u, c_1, z_1)\) and \((u, c_2, z_2)\) are both accepting transcripts. This means that
\[
\begin{align*}
g^{z_1} &= u \cdot h^{c_1} \quad \text{and} \quad g^{z_2} = u \cdot h^{c_2} \\
\Rightarrow \quad \frac{g^{z_1}}{h^{c_1}} &= \frac{g^{z_2}}{h^{c_2}} \quad \Rightarrow \quad g^{z_1 + c_1 c_2} = g^{z_2 + c_2} \quad \text{with overwhelming probability}, \\
\Rightarrow \quad \chi = (z_1 - z_2)(c_1 - c_2)^{-1} \in Z_p \quad \text{and } c_1 \neq c_2.
\end{align*}
\]

Thus, extractor succeeds with overwhelming probability.

\[ \text{(Boneh-Shoup, Lemma 19.2)} \]

If \( P^* \) succeeds with probability \( \epsilon \), then need to rely on "Rewinding Lemma" to argue that extractor obtains two accepting transcripts with probability at least \( \epsilon^2 - \frac{1}{p} \).

The ability to extract a witness from any two accepting transcripts is very useful

\( \rightarrow \) called special soundness (for 3-message protocols)

given \((u, t_1, z_1)\) and \((u, t_2, z_2)\) \( \Rightarrow \) can extract the witness

[ initial message, challenge, response ] [same initial message, different challenges]
3-message protocols that satisfy completeness, special soundness, and HVZK are called I-protocols. I-protocols are useful for building signatures and identification protocols.

How can a prover both prove knowledge and yet be zero-knowledge at the same time?

Extractor operates by "rewinding" the prover. If the prover has good success probability, it can answer most challenges correctly. But in the real (actual) protocol, verifier cannot rewind (i.e., verifier only sees prover on fresh protocol executions), which can provide zero-knowledge.

Many extensions of Schnorr's protocol to prove relations in the exponent.

For example, suppose we want to prove that \((g, h, u, v)\) is a DDH tuple (i.e., \(\exists x \in \mathbb{Z}_p : h^x g^x = v\)).

\[
\begin{array}{c|c|c}
\text{prover} & g^r & u^r \\
\hline
\text{verifier} & t & (r + \alpha t) \\
\end{array}
\]

Check that \(g^{r+\alpha t} = (g^r)^h \cdot (u^r)^{vt}\).

Completeness: Follows by construction.

Special soundness: Suppose we have accepting transcripts \(((a, p), t_1, z_1)\) and \(((a, p), t_2, z_2)\).

Then \(\frac{g^{z_1}}{g^{z_2}} = \frac{a^{t_1}}{a^{t_2}} \Rightarrow z_1 - z_2 = \alpha (t_1 - t_2) \quad (\text{mod } p)\)

\(\Rightarrow \alpha = (z_1 - z_2)(t_1 - t_2)^{-1}\)

Similar calculation shows that \(v = u^\alpha\).

HVZK: Construct simulator as follows:
1. Sample \(z \leftarrow \mathbb{Z}_p\)
2. Sample \(t \leftarrow \mathbb{Z}_p\)
3. Output \((g^z, v^z, t, z)\)

Since \(z\) is uniform and \(h = g^x, v = u^\alpha\), distribution of \((g^z, v^z, t, z)\) is identical to \((g^r, u^r)\). Challenge \(t\) is identically distributed, which uniquely determines \(z\).