Zero-Knowledge Proofs for Classical Planning Problems: Concrete Example

Augusto B. Corrêa, Clemens Büchner, Remo Christen
University of Basel, Switzerland
{augusto.blaascorrea,clemens.buechner,remo.christen}@unibas.ch

Here we give a concrete example for the protocol ZK-BoundedPlanEx of the paper "*Zero-Knowledge Proofs for Classical Planning Problems*" (Corrêa, Büchner, and Christen 2023). We follow the notation used in the paper throughout our example.

Step 0

Both prover P and verifier V have as common input $\langle \Pi, k\rangle$ where $k = 3$ and Π is the planning task $\Pi = \langle \mathcal{V}, \mathcal{A}, I, G\rangle$, with

$$\begin{aligned}\mathcal{V} &= \{v_1, v_2, v_3\},\\ \mathcal{A} &= \{a_1, a_2\},\\ I &= \{\},\\ G &= \{v_1, v_2, v_3\},\end{aligned}$$

and

$$\begin{aligned}\mathit{pre}(a_1) &= \{\},\\ \mathit{eff}(a_1) &= \{v_1\},\\ \mathit{pre}(a_2) &= \{v_1\},\\ \mathit{eff}(a_2) &= \{\neg v_1, v_2, v_3\}.\end{aligned}$$

The plan for Π known by P is $\pi = \langle a_1, a_2, a_1\rangle$.

Step 1

Next we detail the transformations (a)–(e) described in Step 1 of the ZK-BoundedPlanEx.

(a) The first transformation adds a dummy action $a_\varnothing$ to the task, leading to $\Pi_\text{a} = \langle \mathcal{V}_\text{a}, \mathcal{A}_\text{a}, I_\text{a}, G_\text{a}\rangle$ where

$$\begin{aligned}\mathcal{V}_\text{a} &= \mathcal{V} = \{v_1, v_2, v_3\}\\ \mathcal{A}_\text{a} &= \{a_1, a_2, a_\varnothing\}\\ I_\text{a} &= I = \{\}, \text{ and}\\ G_\text{a} &= G = \{v_1, v_2, v_3\}\end{aligned}$$

and

$$\mathit{pre}(a_\varnothing) = \mathit{eff}(a_\varnothing) = \{\}.$$

(b) Let us briefly refresh some definitions from the paper. Recall that

$$\begin{aligned}m(a) &= |\mathit{vars}(\mathit{pre}(a)) \cap \mathit{vars}(\mathit{eff}(a))|, \text{for all } a \in \mathcal{A};\\ m^* &= \max_{a\in\mathcal{A}} m(a).\end{aligned}$$

In our running example, we have $m(a_1) = m(a_\varnothing) = 0$ and $m(a_2) = m^* = 1$ because v_1 occurs in $\mathit{pre}(a_2)$ as well as in $\mathit{eff}(a_2)$, but no other variable occurs in both the precondition and effect of the same action in $\mathcal{A}_\text{a}$. Hence, we introduce two new variables $m_1^{a_1}, m_1^{a_\varnothing}$ and six new actions $a_1^\bot$, $a_1^\top$, $a_2^\bot$, and $a_2^\top$, $a_\varnothing^\bot$, and $a_\varnothing^\bot$ where

$$\begin{aligned}\mathit{pre}(a_1^\bot) &= \{\neg m_1^{a_1}\},\\ \mathit{eff}(a_1^\bot) &= \{v_1, m_1^{a_1}\},\\ \mathit{pre}(a_1^\top) &= \{m_1^{a_1}\},\\ \mathit{eff}(a_1^\top) &= \{v_1, \neg m_1^{a_1}\},\\ \mathit{pre}(a_2^\bot) = \mathit{pre}(a_2^\top) = \mathit{pre}(a_2) &= \{v_1\},\\ \mathit{eff}(a_2^\bot) = \mathit{eff}(a_2^\top) = \mathit{eff}(a_2) &= \{\neg v_1, v_2, v_3\},\\ \mathit{pre}(a_\varnothing^\bot) &= \{\neg m_1^{a_\varnothing}\},\\ \mathit{eff}(a_\varnothing^\bot) &= \{m_1^{a_\varnothing}\},\\ \mathit{pre}(a_\varnothing^\top) &= \{m_1^{a_\varnothing}\}, \text{ and}\\ \mathit{eff}(a_\varnothing^\top) &= \{\neg m_1^{a_\varnothing}\}.\end{aligned}$$

Note that this part of the transformation does not affect the initial state and the goal. As a result, we obtain $\Pi_\text{b} = \langle \mathcal{V}_\text{b}, \mathcal{A}_\text{b}, I_\text{b}, G_\text{b}\rangle$ where

$$\begin{aligned}\mathcal{V}_\text{b} &= \{v_1, v_2, v_3, m_1^{a_\varnothing}, m_1^{a_1}\},\\ \mathcal{A}_\text{b} &= \{a_1^\bot, a_1^\top, a_2^\bot, a_2^\top, a_\varnothing^\bot, a_\varnothing^\top\},\\ I_\text{b} &= I = \{\}, \text{ and}\\ G_\text{b} &= G = \{v_1, v_2, v_3\}\end{aligned}$$

Observe that $\pi_\text{b} = \langle a_1^\bot, a_2^\bot, a_1^\top\rangle$ is a plan for Π_b that corresponds to π in Π. This is done by simply replacing every i-th occurrence of a in π with $a^\bot$ if i is odd and with $a^\top$ if i is even.

(c) In the paper, p and e were defined as follows

$$\begin{aligned}p(a) &= |\mathit{pre}(a)|,\\ e(a) &= |\mathit{eff}(a)|\end{aligned}$$

for all $a \in \mathcal{A}$, and

$$\begin{aligned}p^* &= \max_{a\in\mathcal{A}'} p(a),\\ e^* &= \max_{a\in\mathcal{A}'} e(a).\end{aligned}$$

In Π_b, we have $p^* = 1$ because all actions in $\mathcal{A}_\mathrm{b}$ have exactly one precondition and $e^* = 3$ because $a_2^\perp$ and $a_2^\top$ have the most effects, namely 3. As $a_\varnothing^\perp$ and $a_\varnothing^\top$ have only one variable in their effects and $a_1^\perp$ and $a_1^\top$ have only two variables in their effects, we need to add two (respectively one) additional variable(s) for each of them, respectively. Hence, we need to introduce six new variables $e_1^{a_1^\perp}$, $e_1^{a_1^\top}$ $e_1^{a_\varnothing^\perp}$, $e_1^{a_\varnothing^\top}$ $e_2^{a_\varnothing^\perp}$, and $e_2^{a_\varnothing^\top}$. Furthermore, we define the new actions $\overline{a}_1^\perp$, $\overline{a}_1^\top$, $\overline{a}_2^\perp$, $\overline{a}_2^\top$, $\overline{a}_\varnothing^\perp$, and $\overline{a}_\varnothing^\top$ where

$$\begin{aligned}
\mathit{pre}(\overline{a}_1^\perp) &= \mathit{pre}(a_1^\perp) = \{\neg m_1^{a_1}\},\\
\mathit{eff}(\overline{a}_1^\perp) &= \{v_1, m_1^{a_1}, \neg e_1^{a_1^\perp}\},\\
\mathit{pre}(\overline{a}_1^\top) &= \mathit{pre}(a_1^\top) = \{m_1^{a_1}\},\\
\mathit{eff}(\overline{a}_1^\top) &= \{v_1, \neg m_1^{a_1}, \neg e_1^{a_1^\top}\},\\
\mathit{pre}(\overline{a}_2^\perp) = \mathit{pre}(\overline{a}_2^\top) &= \mathit{pre}(a_2) = \{v_1\},\\
\mathit{eff}(\overline{a}_2^\perp) = \mathit{eff}(\overline{a}_2^\top) &= \mathit{eff}(a_2) = \{\neg v_1, v_2, v_3\},\\
\mathit{pre}(\overline{a}_\varnothing^\perp) &= \mathit{pre}(a_\varnothing^\perp) = \{\neg m_1^{a_\varnothing}\},\\
\mathit{eff}(\overline{a}_\varnothing^\perp) &= \{m_1^{a_\varnothing}, \neg e_1^{a_\varnothing^\perp}, \neg e_2^{a_\varnothing^\perp}\},\\
\mathit{pre}(\overline{a}_\varnothing^\top) &= \mathit{pre}(a_\varnothing^\top) = \{m_1^{a_\varnothing}\}, \text{ and}\\
\mathit{eff}(\overline{a}_\varnothing^\top) &= \{\neg m_1^{a_\varnothing}, \neg e_1^{a_\varnothing^\top}, \neg e_2^{a_\varnothing^\top}\}.
\end{aligned}$$

As a result, we obtain $\Pi_\mathrm{c} = \langle \mathcal{V}_\mathrm{c}, \mathcal{A}_\mathrm{c}, I_\mathrm{c}, G_\mathrm{c}\rangle$ where

$$\begin{aligned}
\mathcal{V}_\mathrm{c} &= \{v_1, v_2, v_3, m_1^{a_1}, m_1^{a_\varnothing},\\
&\quad e_1^{a_1^\perp}, e_1^{a_1^\top}, e_1^{a_\varnothing^\perp}, e_1^{a_\varnothing^\top}, e_2^{a_\varnothing^\perp}, e_2^{a_\varnothing^\top}\},\\
\mathcal{A}_\mathrm{c} &= \{\overline{a}_1^\perp, \overline{a}_1^\top, \overline{a}_2^\perp, \overline{a}_2^\top, \overline{a}_\varnothing^\perp, \overline{a}_\varnothing^\top\},\\
I_\mathrm{c} &= I = \{\}, \text{ and}\\
G_\mathrm{c} &= G = \{v_1, v_2, v_3\}
\end{aligned}$$

The plan π_b is changed by simply using the corresponding new actions. Thus, the plan for Π_c is $\pi_\mathrm{c} = \langle \overline{a}_1^\perp, \overline{a}_2^\perp, \overline{a}_1^\top\rangle$, which corresponds to π_b in Π_b.

(d) Recall that ρ is a function permuting variables labels and swapping the truth values of a subset of variables that is chosen uniformly at random. These swaps are done consistently across the entire task. To simplify our example, we will use ρ as a function mapping $\mathcal{V}$ to some alphabet $\{\rho_1, \dots, \rho_{11}\}$ although it should be a permutation function. Also to make the example easier, we modify the names given to the actions. Note that *a priori* action names are simply syntactic sugar as they are simply identified by their preconditions and effects, which will already be permuted by ρ.

Assume ρ is such that the variables are renamed as follows:

$$\begin{aligned}
\rho_1 &:= v_3, & a_1^\rho &:= \overline{a}_1^\top,\\
\rho_2 &:= e_1^{a_1^\perp}, & a_2^\rho &:= \overline{a}_2^\top,\\
\rho_3 &:= \neg v_2, & a_3^\rho &:= \overline{a}_2^\perp,\\
\rho_4 &:= m_1^{a_1}, & a_4^\rho &:= \overline{a}_1^\perp,\\
\rho_5 &:= v_1, & a_5^\rho &:= \overline{a}_\varnothing^\top, \text{ and}\\
\rho_6 &:= \neg e_1^{a_1^\top} & a_6^\rho &:= \overline{a}_\varnothing^\perp.\\
\rho_7 &:= \neg e_2^{a_\varnothing^\perp},\\
\rho_8 &:= e_1^{a_\varnothing^\top},\\
\rho_9 &:= \neg m_1^{a_\varnothing},\\
\rho_{10} &:= e_1^{a_\varnothing^\perp}, \text{ and}\\
\rho_{11} &:= e_2^{a_\varnothing^\top};
\end{aligned}$$

(If we write $\rho := \neg v$ we mean ρ corresponds to the old variable v but its value in I_c is inverted.) Then $\Pi_\mathrm{d} = \langle \mathcal{V}_\mathrm{d}, \mathcal{A}_\mathrm{d}, I_\mathrm{d}, G_\mathrm{d}\rangle$ looks as follows:

$$\begin{aligned}
\mathcal{V}_\mathrm{d} &= \{\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \rho_7, \rho_8, \rho_9, \rho_{10}, \rho_{11}\},\\
\mathcal{A}_\mathrm{d} &= \{a_1^\rho, a_2^\rho, a_3^\rho, a_4^\rho, a_5^\rho, a_6^\rho\},\\
I_\mathrm{d} &= \{\rho_3, \rho_6, \rho_7, \rho_9\}, \text{ and}\\
G_\mathrm{d} &= \{\rho_1, \neg\rho_3, \rho_5\}
\end{aligned}$$

where

$$\begin{aligned}
\mathit{pre}(a_1^\rho) &= \{\rho_4\},\\
\mathit{eff}(a_1^\rho) &= \{\neg\rho_4, \rho_5, \rho_6\},\\
\mathit{pre}(a_2^\rho) &= \{\rho_5\},\\
\mathit{eff}(a_2^\rho) &= \{\rho_1, \neg\rho_3, \neg\rho_5\},\\
\mathit{pre}(a_3^\rho) &= \{\rho_5\},\\
\mathit{eff}(a_3^\rho) &= \{\rho_1, \neg\rho_3, \neg\rho_5\},\\
\mathit{pre}(a_4^\rho) &= \{\neg\rho_4\},\\
\mathit{eff}(a_4^\rho) &= \{\neg\rho_2, \rho_4, \rho_5\},\\
\mathit{pre}(a_5^\rho) &= \{\neg\rho_9\},\\
\mathit{eff}(a_5^\rho) &= \{\neg\rho_8, \rho_9, \neg\rho_{11}\},\\
\mathit{pre}(a_6^\rho) &= \{\rho_9\}, \text{ and}\\
\mathit{eff}(a_6^\rho) &= \{\rho_7, \neg\rho_9, \neg\rho_{10}\}.
\end{aligned}$$

Observe that $\pi_\mathrm{d} = \langle a_4^\rho, a_3^\rho, a_1^\rho\rangle$ is the plan for Π_d that corresponds to π_c in Π_c.

(e) After introducing the artificial initial and goal state as well as the initial and goal actions, we obtain $\Pi_\mathrm{e} = \langle \mathcal{V}_\mathrm{e}, \mathcal{A}_\mathrm{e}, I_\mathrm{e}, G_\mathrm{e}\rangle$ where

$$\begin{aligned}
\mathcal{V}_\mathrm{e} &= \{\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \rho_7, \rho_8, \rho_9, \rho_{10}, \rho_{11}, v_I, v_*\},\\
\mathcal{A}_\mathrm{e} &= \{\hat{a}_1^\rho, \hat{a}_2^\rho, \hat{a}_3^\rho, \hat{a}_4^\rho, \hat{a}_5^\rho, \hat{a}_6^\rho, a_I, a_*\},\\
I_\mathrm{e} &= \{v_I\}, \text{ and}\\
G_\mathrm{e} &= \{\neg\rho_1, \dots, \neg\rho_{11}, \neg v_I, v_*\}
\end{aligned}$$

where

$$
\begin{aligned}
\mathit{pre}(\hat{a}_1^\rho) &= \{\rho_4, \neg v_I\},\\
\mathit{eff}(\hat{a}_1^\rho) &= \{\neg\rho_4, \rho_5, \rho_6\},\\
\mathit{pre}(\hat{a}_2^\rho) &= \{\rho_5, \neg v_I\},\\
\mathit{eff}(\hat{a}_2^\rho) &= \{\rho_1, \neg\rho_3, \neg\rho_5\},\\
\mathit{pre}(\hat{a}_3^\rho) &= \{\rho_5, \neg v_I\},\\
\mathit{eff}(\hat{a}_3^\rho) &= \{\rho_1, \neg\rho_3, \neg\rho_5\},\\
\mathit{pre}(\hat{a}_4^\rho) &= \{\neg\rho_4, \neg v_I\},\\
\mathit{eff}(\hat{a}_4^\rho) &= \{\neg\rho_2, \rho_4, \rho_5\},\\
\mathit{pre}(\hat{a_5}^\rho) &= \{\neg\rho_9, \neg v_I\},\\
\mathit{eff}(\hat{a_5}^\rho) &= \{\neg\rho_8, \rho_9, \neg\rho_{11}\},\\
\mathit{pre}(\hat{a_6}^\rho) &= \{\rho_9, \neg v_I\},\\
\mathit{eff}(\hat{a_6}^\rho) &= \{\neg\rho_7, \neg\rho_9, \rho_{10}\}.\\
\mathit{pre}(a_I) &= \{v_I\},\\
\mathit{eff}(a_I) &= \{\rho_3, \rho_6, \rho_7, \rho_9, \neg v_I\},\\
\mathit{pre}(a_*) &= \{\rho_1, \neg\rho_3, \rho_5, \neg v_I\}, \text{ and}\\
\mathit{eff}(a_*) &= \{\neg\rho_1, \ldots, \neg\rho_{11}, \neg v_I, v_*\}.
\end{aligned}
$$

Finally, note that we can transform the plan π_d into a valid plan for Π_e simply by appending a_I and a_* to the plan and switching each action to its new corresponding one. Thus, $\pi_\mathrm{e} = \langle a_I, \hat{a}_4^\rho, \hat{a}_3^\rho, \hat{a}_1^\rho, a_* \rangle$ is the plan for Π_e that corresponds to π_d in Π_d.

Given that the entire task transformation is called $\hat{\Pi} = \Pi_\mathrm{e}$ in the main paper, we also refer to π_e as $\hat{\pi}$.

Step 2

The prover P then creates the sequence of states $S = \langle s_0, s_1, s_2, s_3, s_4, s_5 \rangle$ where

$$
\begin{aligned}
s_0 &= \{v_I\}\\
s_1 &= \{\rho_3, \rho_6, \rho_7, \rho_9\}\\
s_2 &= \{\rho_3, \rho_4, \rho_5, \rho_6, \rho_7, \rho_9\}\\
s_3 &= \{\rho_1, \rho_4, \rho_6, \rho_7, \rho_9\}\\
s_4 &= \{\rho_1, \rho_5, \rho_6, \rho_7, \rho_9\}\\
s_5 &= \{v_*\}.
\end{aligned}
$$

P sends *Commit*$(\hat{\Pi})$, *Commit*$(\hat{\pi})$, and *Commit*(S) to V.

Step 3

Recall that $\ell = k + 2 = 5$ in our case, as $k = 3$. V checks that $|\mathit{Commit}(\hat{\pi})| > \ell = 5$. This is not the case, so V continues the protocol: it picks a random bit $b \in \{0, 1\}$ and sends it to P.

If $b = 0$, the case is trivial: P sends the function σ used to produce $\hat{\Pi}$ and V checks that $\sigma(\Pi) = \hat{\Pi}$. In our example, this is true so V would accept the protocol.

Let us continue the protocol assuming that $b = 1$. Then P opens s_0 and s_5 from *Commit*(S).

Step 4

V checks if the opened states are what it expects: s_0 should be the initial state $\{v_I\}$ as defined by the protocol; s_5 should be the unique goal state $\{v_*\}$ as defined by the protocol. V verifies that this is indeed the case.

The protocol continues: V now uniformly chooses an integer $m \in \{1, 2, 3, 4, 5\}$ and sends it to P. In our example, let us assume that V picked $m = 3$ – although any choice would lead to the same protocol conclusion. This means that the verifier will check the third transition of the plan, denoted as $(s_2, \hat{a}_3^\rho, s_3)$.

Step 5

P opens $\hat{\mathcal{V}}$ from *Commit*$(\hat{\Pi})$, and reveals $\hat{a}_3^\rho$ from *Commit*$(\hat{\mathcal{A}})$. As $m = 3$, P also opens s_2 and s_3 from *Commit*(S), as well as $\hat{a}_3^\rho$ from *Commit*$(\hat{\pi})$. Note that P reveals $\hat{a}_3^\rho$ twice: once in $\hat{\mathcal{A}}$ and once in $\hat{\pi}$. This is done so the verifier can check that the action in the transition is indeed in the transformed task description.

V then checks that all variables used in $s_2, \hat{a}_3^\rho$ and s_3 are indeed in $\hat{\mathcal{V}}$. This is clearly the case in our example.

It also compares the action $\hat{a}_3^\rho$ obtained from *Commit*$(\hat{\mathcal{A}})$ and the action obtained from the third transition of *Commit*$(\hat{\pi})$. Both of them are indeed $\hat{a}_3^\rho$. Last, V computes $s_2[\![\hat{a}_3^\rho]\!]$. This is also true, so V finally accepts the protocol.

References

Corrêa, A. B.; Büchner, C.; and Christen, R. 2023. Zero-Knowledge Proofs for Classical Planning Problems. In Chen, Y.; and Neville, J., eds., *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI 2023)*. AAAI Press.