Explaining Hyperproperty Violations

Norine Coenen¹, Raimund Dachselt², Bernd Finkbeiner¹, Hadar Frenkel¹, Christopher Hahn¹, Tom Horak², Niklas Metzger¹, Julian Siber¹

¹CISPA Helmholtz Center for Information Security
²Technische Universität Dresden

As presented at the 34th International Conference on Computer-Aided Verification (CAV 2022).
Model Checking

System model

Specification

Model Checker

✅

✗

⚡️
Explaining Counterexamples

We give explanations by identifying *causes* in the non-deterministic input sequences.

E.g.: *Explaining Counterexamples Using Causality*. Beer, Ben-David, Chockler, Orni, and Trefler. (CAV 2009).
*Error explanation with distance metrics*. Groce, Chaki, Kroening, Strichman. Int. J. Softw. Tools Technol. Transf. 8 (2006)
Hyperproperties

Observational determinism: “A system appears deterministic to low-security users”.

(Generalized) Noninterference

Declassification

Trace equality

Hyperproperties. Clarkson and Schneider. (CSF 2008).
HyperLTL

Observational determinism: “A system appears deterministic to low-security users”.

$$\forall \pi. \forall \pi'. \Box (li_\pi \leftrightarrow li_{\pi'}) \rightarrow \Box (lo_\pi \leftrightarrow lo_{\pi'})$$

(Generalized) Noninterference

Declassification

Trace equality

Temporal Logics for Hyperproperties. Clarkson, Finkbeiner, Koleini, Micinski, Rabe, and Sánchez. (POST 2014).
HyperLTL Model Checking

**Specification \( \varphi \):**

\[
\forall \pi. \forall \pi'. \Box (li_\pi \leftrightarrow li_{\pi'}) \rightarrow \Box (lo_\pi \leftrightarrow lo_{\pi'})
\]

**Counterexample \( \Gamma \):**

\[
\begin{array}{c|c|c}
\pi & \{} & \{} \\
\pi' & \{hi\} & \{hi, lo\} \\
\end{array}
\]

**System \( T \):**

![System Diagram]

Violation of \( \varphi \) on \( \Gamma \) is due to interactions between inputs on multiple traces.

*Temporal Logics for Hyperproperties.* Clarkson, Finkbeiner, Koleini, Micinski, Rabe, and Sánchez. (POST 2014).

*Algorithms for Model Checking HyperLTL and HyperCTL*. Finkbeiner, Rabe, and Sánchez. (CAV 2015).
HyperLTL Model Checking

Specification $\varphi$:

$$\forall \pi. \forall \pi'. \square (\text{hi}_\pi \leftrightarrow \text{hi}_{\pi'}) \rightarrow \square (\text{lo}_\pi \leftrightarrow \text{lo}_{\pi'})$$

Counterexample $\Gamma$:

$$\pi = \begin{cases} \{\} & \{\} & \{\} \omega \\
\{\text{hi}\} & \{\text{hi}, \text{lo}\} & \{\} \omega 
\end{cases}$$

Violation of $\varphi$ on $\Gamma$ is due to interactions between inputs on multiple traces.

Temporal Logics for Hyperproperties. Clarkson, Finkbeiner, Koleini, Micinski, Rabe, and Sánchez. (POST 2014).

Algorithms for Model Checking HyperLTL and HyperCTL*. Finkbeiner, Rabe, and Sánchez. (CAV 2015).
Causal Analysis

Specification $\varphi$:

$$\forall \pi. \forall \pi'. \, \square (li_\pi \leftrightarrow li_{\pi'}) \rightarrow \square (lo_\pi \leftrightarrow lo_{\pi'})$$

Counterexample $\Gamma$:

$$\begin{array}{c|c|c}
\pi & \emptyset & \emptyset \\
\pi' & \{hi\} & \{hi, lo\} \\
& \emptyset & \emptyset^\omega \\
\hline
\end{array}$$

System $T$:

![Diagram](image)

We highlight the *causes* on the input sequences.

*Causes and Explanations: A Structural-Model Approach.* Halpern and Pearl. Brit. J. Phil. Sci. 56 (2005).
*A Modification of the Halpern-Pearl Definition of Causality.* Halpern. (IJCAI 2015).
We extend HP’s actual causality to hyperproperty effects and reactive systems.
Events and Causes

\[
\begin{array}{c|c|c}
\pi &= & \{\} \quad \{\} \quad \{\}^\omega \\
\pi' &=& \{hi\} \quad \{hi, lo\} \quad \{\}^\omega \\
\end{array}
\]

An event \(\langle l_a, n, \pi \rangle\) is the value of an atomic proposition \(a\) at position \(n\) in \(\pi\).

\[
(\pi, \pi') \models \langle hi, 0, \pi' \rangle
\]
Events and Causes

\[
\pi = \begin{array}{c|c|c}
\{\} & \{\} & \{\}^\omega \\
\hline
\{hi\} & \{hi, lo\} & \{\}^\omega
\end{array}
\]

An event \(\langle l_a, n, \pi \rangle\) is the value of an atomic proposition \(a\) at position \(n\) in \(\pi\).

A cause \(C\) is a set of events.
$C$ is a Cause if...

\[
\begin{align*}
\pi &= \{\} & \{\} & \{\}^\omega \\
\pi' &= \{hi\} & \{hi, lo\} & \{\}^\omega
\end{align*}
\]

**SAT:** $\Gamma$ satisfies all events in $C$. 
Interventions

$$\pi = \{\} \quad \{\} \quad \{\}^\omega$$

$$\pi' = \{hi\} \quad \{hi, lo\} \quad \{\}^\omega$$

An intervention on $C$ flips the values of all events in $C$.

$$\text{intervene}(\Gamma, \{\langle hi, 0, \pi'\rangle\}, \emptyset) \quad \rightarrow \quad \pi = \{\} \quad \{\} \quad \{\}^\omega$$

$$\pi' = \{\} \quad \{hi\} \quad \{lo\} \quad \{\}^\omega$$
Contingencies

\[ \pi = \{\} \quad \{\} \quad \{\}^\omega \]

\[ \pi' = \{hi\} \quad \{hi, lo\} \quad \{\}^\omega \]

A **contingency** \( \mathcal{W} \) allows to reset states back to \( \Gamma \).

\[ \text{intervene}(\Gamma, \{\langle hi, 0, \pi'\rangle\}, \{\langle \neg lo, 2, \pi'\rangle\}) \]

\[ \pi = \{\} \quad \{\} \quad \{\} \quad \{\}^\omega \]

\[ \pi' = \{\} \quad \{hi\} \quad \{\} \quad \{\}^\omega \]
$C$ is a Cause if...

\[
\begin{array}{c|c|c}
\pi &=& \{\} \quad \{\} \quad \{\}^\omega \\
\pi' &=& \{hi\} \quad \{hi, lo\} \quad \{\}^\omega \\
\end{array}
\]

**SAT:** $\Gamma$ satisfies all events in $C$.

**CF:** There exists a $\mathcal{W}$ and $C' \subseteq C$ s.t.: $\text{intervene}(\Gamma, C', \mathcal{W}) \models \varphi$.

**MIN:** No $C' \subset C$ satisfies SAT and CF.
**$C$ is a Cause if...**

\[ \pi = \begin{cases} \{\} & \{} \\ \{hi\} & \{hi, lo\} \end{cases} \quad \{\}^\omega \]

\[ \pi' = \begin{cases} \{\} & \{} \\ \{hi\} & \{hi, lo\} \end{cases} \quad \{\}^\omega \]

**SAT:** $\Gamma$ satisfies all events in $C$.

**CF:** There exists a $\mathcal{W}$ and $C' \subseteq C$ s.t.: $\text{intervene}(\Gamma, C', \mathcal{W}) \models \varphi$.

**MIN:** No $C' \subset C$ satisfies SAT and CF.
Encoding the CF Criterion

System $T$:

(Partial) counterexample:

$$\pi' = \{hi\}\{hi, lo\}\{}^\omega$$

Counterfactual automaton $(T, \pi')$:

Counterfactual automata have additional inputs (here: $c$) for setting a contingency.
Finding a Cause as a Hyperproperty

\[ \pi = \{\} \quad \{\} \quad \{\}^\omega \]
\[ \pi' = \{hi\} \quad \{hi, lo\} \quad \{\}^\omega \]

Counterexample

Counterfactual Automata

\[ \exists \pi_c. \exists \pi'_c. \forall \pi''_c. \forall \pi'''_c. \varphi_{cause} \]

Causality

\[ \forall \pi. \forall \pi'. (\square (li_\pi \leftrightarrow li_{\pi'}) \rightarrow \square (lo_\pi \leftrightarrow lo_{\pi'}) ) \]

HyperLTL Specification

Encoding of causality in \( \varphi_{cause} \): see our paper.
Computing All Causes

If some $C$ is a cause, then no strict superset $C' \supset C$ is a cause.
## Experiments

| Instance              | $|\Gamma|$ | $|\varphi|$ | $(C)$ | time (ms) |
|-----------------------|-----------|-------------|-------|-----------|
| Running example (paper)| 10        | 9           | 2     | 55        |
| Security in & out     | 35        | 19          | 8     | 798       |
| Drone example 1       | 24        | 19          | 5     | 367       |
| Drone example 2       | 18        | 36          | 3     | 256       |
| Asymmetric arbiter ’19| 28        | 35          | 10    | 490       |
| Asymmetric arbiter    | 72        | 35          | 24    | 1480      |
Visual Analysis of Hyperproperties for Understanding Model Checking Results. Horak, Coenen, Metzger, Hahn, Flemisch, Méndez, Dimov, Finkbeiner, and Dachselt. (VIS 2021). https://hypervis.tools.react.cs.uni-saarland.de
Conclusion

Counterexamples of hyperproperties are *difficult to understand* and *debug*.

Extending HP’s actual causality to hyperproperties gives *precise explanations*.

*Causal inference* can itself be stated as a *hyperproperty* model-checking problem.

Symbolic causes, explicit relations, existential quantifiers