Topological Semantics for Common Inductive Knowledge

Solving an inductive version of the coordinated attack problem

Siddharth Namachivayam

Blockchains/Web3

- **Many on-chain applications rely on attestations from validators about off-chain data.**
- Example: Proof of Reserve for freezing wrapped coins whenever backing drops.
- **Consensus from validators drives on-chain actions.**
- Example: Validators must agree that there is sufficient collateral to unfreeze a wrapped coin.
- **False positives are often far worse than false negatives.**
- Example: Unfreezing a wrapped coin with insufficient collateral is worse than freezing the coin when there is sufficient collateral.

Inductive Coordinated Attack

- Validators passively receive signals about backing and follow an exogenous asynchronous communication protocol.
- They each tolerate a fixed number of mind switches about whether there is eventually always (e.a.) sufficient collateral.
- Want to find an attestation protocol which satisfies:
    1. **Validity-** If backing always eventually (a.e.) drops then not a single validator converges on attesting that there is e.a. sufficient collateral.
    2. **Agreement-** Either all validators converge on attesting there is e.a. sufficient collateral or no validators converge on attesting there is e.a. sufficient collateral.
    3. **Nontriviality-** At least one execution where all validators converge on attesting there is e.a. sufficient collateral.
- <u>If validators honestly follow such an attestation protocol we can (e.g. via majority vote) ensure a wrapped coin is eventually unfrozen forever whenever all validators correctly converge on attesting there is e.a. sufficient collateral.</u>
- Moreover, we can increase/decrease threshold from simple majority to trade off false positives/false negatives and protect against greater/fewer dishonest validators who deviate from the attestation protocol.

Notation

- N is a set of agents and Ω is a set of possible worlds.
- $\mathscr{E}_i$ is a topological basis over Ω, representing possible information states of agent $i \in N$.
- n_i denotes the maximum number of times agent i is willing to switch their mind after first saying Yes while limit deciding a proposition.
- $\mathscr{T}_i$ is the topology generated by $\mathscr{E}_i$.
- $\mathscr{E}_{i|w}$ denotes those information states in $\mathscr{E}_i$ which contain w.

Strategies

- Each agent i's *strategy* s_i is an n_i-switching method m_i where:

  1. A *method* for agent i is a map $m_i : \mathscr{E}_i \to \{\text{Yes}, ?\}$.

  2. A *t-switching* sequence for m_i is a finite downward sequence $E_0 \supseteq \ldots \supseteq E_t$ of information states in $\mathscr{E}_i$ such that $m_i(E_{2k}) = \text{Yes}$ and $m_i(E_{2k+1}) = ?$.

  3. An *n-switching method* for i has no t-switching sequences for any $t > n$.

- We let $\sigma_{s_i}(w)$ denote the output s_i converges to in world w.

Protocols

- A protocol $(s_i)_{i\in N}$ *satisfies validity for the proposition* P if $\forall i \in N$ we have:

$$\sigma_{s_i}^{-1}(\text{Yes}) \subseteq P.$$

- Next, a protocol $(s_i)_{i\in N}$ *satisfies agreement* if $\forall i, j \in N$ we have:

$$\sigma_{s_i}^{-1}(\text{Yes}) = \sigma_{s_j}^{-1}(\text{Yes}).$$

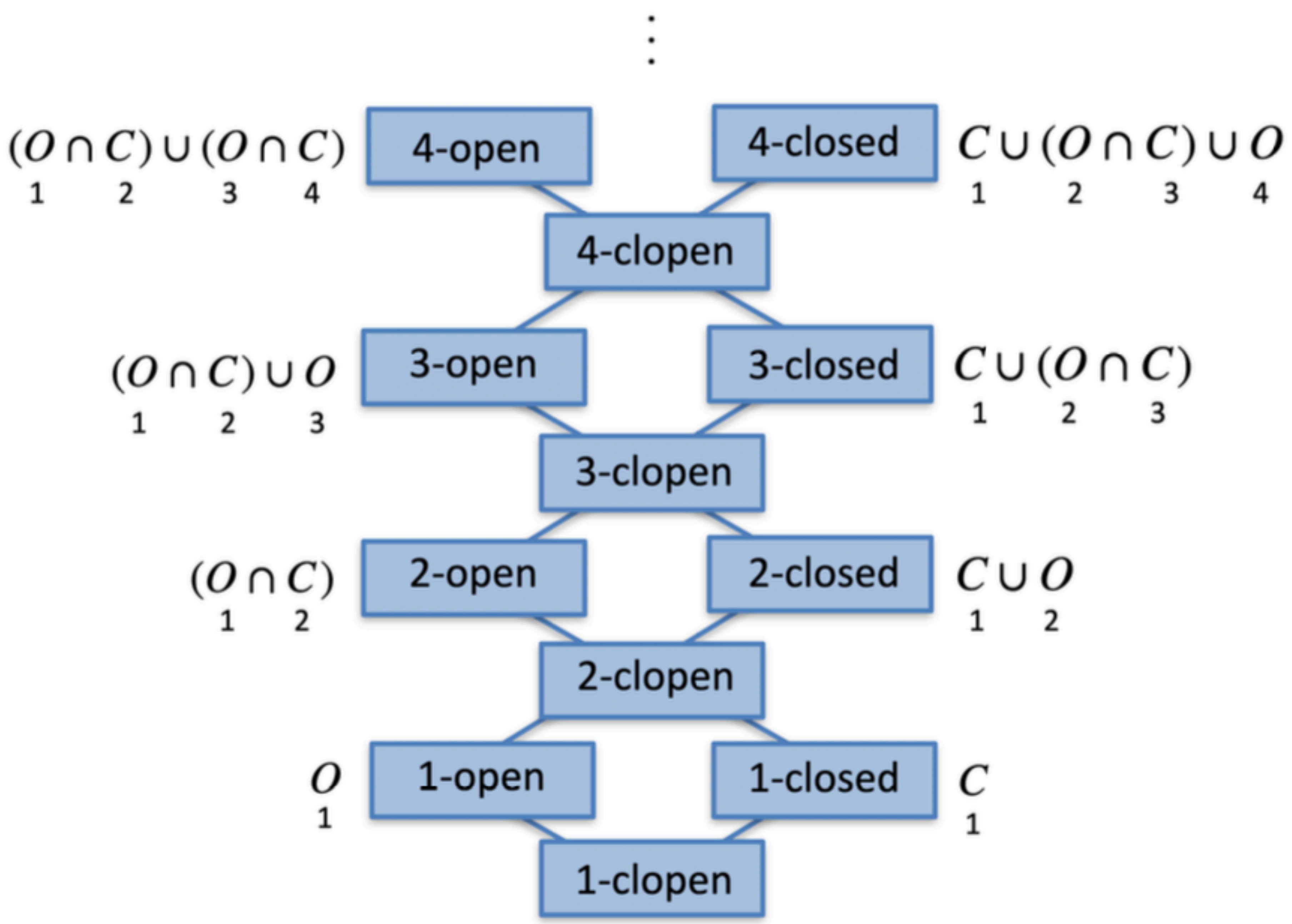
- Finally, a protocol $(s_i)_{i\in N}$ *satisfies nontriviality* if we have:

$$\bigcap_{i\in N} \sigma_{s_i}^{-1}(\text{Yes}) \neq \emptyset.$$

- A protocol $(s_i)_{i\in N}$ satisfying the above conditions *solves consensus for* P.

Topological Difference Hierarchy

- Fact: There exists a strategy s_i where $\sigma_{s_i}^{-1}(\text{Yes}) = W$ iff W is $(n_i + 1)$-open in $\mathscr{T}_i$.
- If W is $(n_i + 1)$-open in $\mathscr{T}_i$ it is easy to reconstruct the strategies s_i such that $\sigma_{s_i}^{-1}(\text{Yes}) = W$.
- Hence, if we can find the non-empty subsets of P are which are $(n_i + 1)$-open in each $\mathscr{T}_i$ then we can reconstruct the protocols which solve consensus for P.
- But how do we find these subsets?



Topological Semantics

- $\text{Yes}_i^{n_i}(W|E)$: i says Yes while limit deciding W in light of evidence E.
- $\mathbb{R}_i W$: i has reason simpliciter to believe W.
- $\mathbb{I}_{i@W}P$: W indicates to i that P.
- $\mathbb{B}_{i@W}P$: i has W as reason to believe P.
- $\mathbb{E}_W P$: everyone has W as reason to believe P.
- $\mathbb{G}_W P$: W generates common inductive knowledge of P.
- $\mathbb{C}P$: P is common inductive knowledge.

- $w \in \mathbb{R}_i W$ iff $\exists E \in \mathscr{E}_{i|w}, \text{Yes}_i^{n_i}(W|E)$.
- $w \in \mathbb{I}_{i@W}P$ iff $\forall E \in \mathscr{E}_{i|w}, \text{Yes}_i^{n_i}(W|E) \to W \cap E \subseteq P$.
- $w \in \mathbb{B}_{i@W}P$ iff $w \in \mathbb{R}_i W \cap \mathbb{I}_{i@W}P$.
- $w \in \mathbb{E}_W P$ iff $w \in \cap_{i \in N} \mathbb{B}_{i@W}P$.
- $w \in \mathbb{E}_W^1 P$ iff $w \in \mathbb{E}_W P$.
- $w \in \mathbb{E}_W^n P$ iff $w \in \mathbb{E}_W \mathbb{E}_W^{n-1} P$.
- $w \in \mathbb{G}_W P$ iff $w \in \cap_{i \in \mathbb{N}^+} \mathbb{E}_W^n P$.
- $w \in \mathbb{C}P$ iff $\exists W \subseteq \Omega, w \in W \cap \mathbb{G}_W P$

Results

- **Theorem 1:** W is a subset of P which is $(n_i + 1)$-open in each $\mathscr{T}_i$ iff it is a fixed point of the map $W \mapsto W \cap \mathbb{G}_W P$. Further, $\forall W \subseteq \Omega$, $W \cap \mathbb{G}_W P$ is itself such a fixed point. This means:

  A. The map $W \mapsto W \cap \mathbb{G}_W P$ can be used to select any subset of P which is $(n_i + 1)$-open in each $\mathscr{T}_i$.

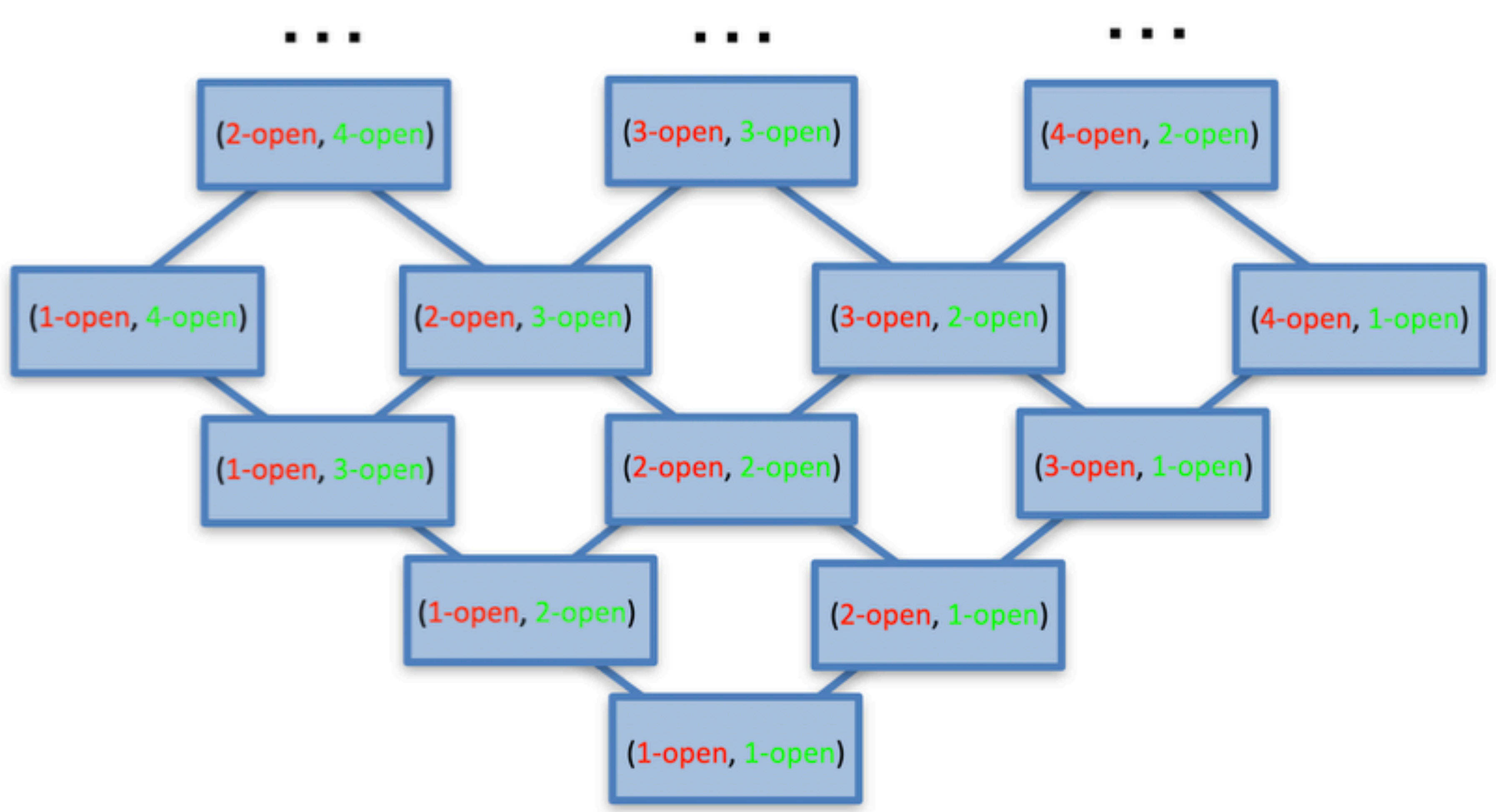
  B. $w \in \mathbb{C}P$ iff w is a world where agents can converge to attesting Yes in some protocol which solves consensus for P.

Results

- **Theorem 2:** $\mathbb{C}P$ does not depend on agents' switching tolerances so long as each $n_i > 0$. To be clear, if agents increase their switching tolerances:

  A. Agents might converge to attesting Yes on a new set in some protocol which solves consensus for P.

  B. However, there will be no new worlds in any of these sets.



Future Work

- We are creating a logical language to reason precisely about common inductive knowledge.
  - Need to develop a sound and complete proof system for our semantics.
  - Will help prove and verify new theorems about our language.
- What kinds of consensus tasks are just inductive coordinated attack problems in disguise?
  - Might want to retemporalize possible worlds semantics back to states of affairs semantics (runs and systems).