Making Geo-Replicated Systems Fast as Possible, Consistent when Necessary

Cheng Li, Daniel Porto, Allen Clement, Johannes Gehrke, Nuno Preguica, Rodrigo Rodrigues

Presented by Sophie Qiu

Motivation:

- Geo-replicated system
- users are globally distributed
- Applications replicate data across datacenters
- Reduce network latencies to users

Dilemma:

- Cross-site consistency latency

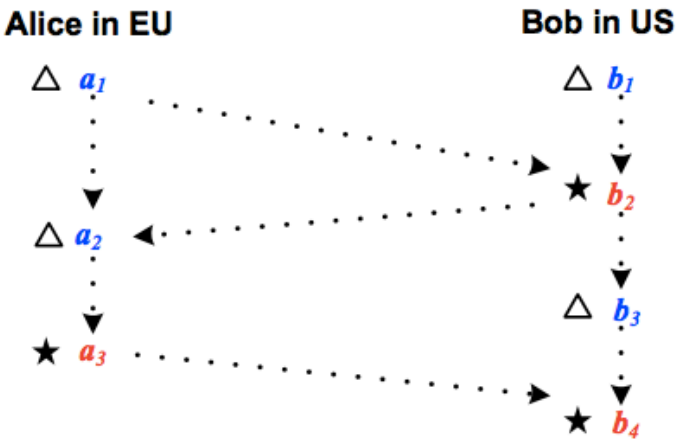
Observation:

- Strong consistency is not always required

Goal:

- RedBlue Consistency: Mixing strong consistency (for application semantics) & eventual consistency (for fast responses) in a same system

Divide Operations into Red and Blue



(a) RedBlue order O of operations

(b) Serializations of O

RedBlue Order:

- Red operations must be totally ordered
- The order of Blue operations can vary from site to site

RedBlue Consistency

Causal serialization

A site has a causal serialization of the RedBlue order if the ordering is a linear extension of the RedBlue order

RedBlue Consistency

Each site applies operations according to the causal serialization of the RedBlue order

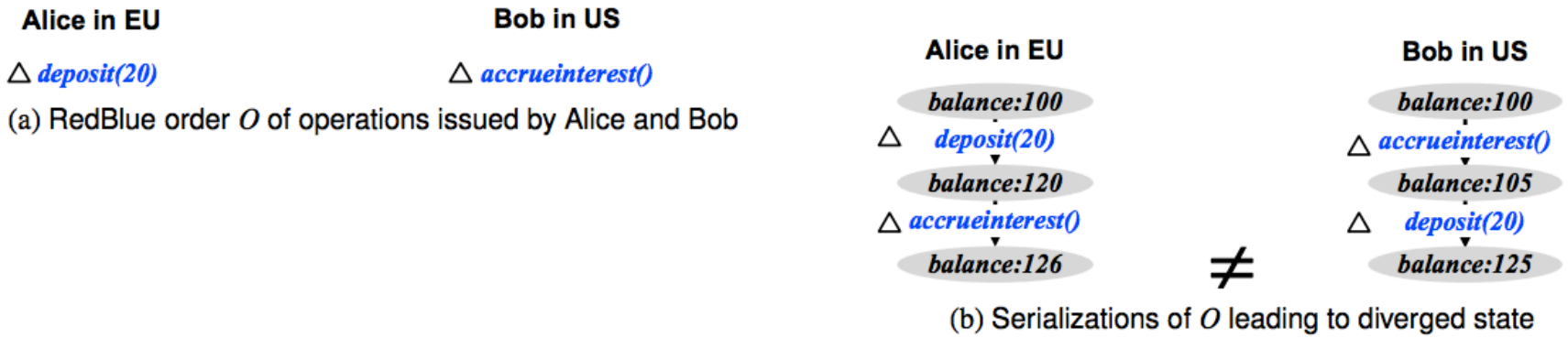


Figure 2: A RedBlue consistent account with initial balance of 100 and final diverged state

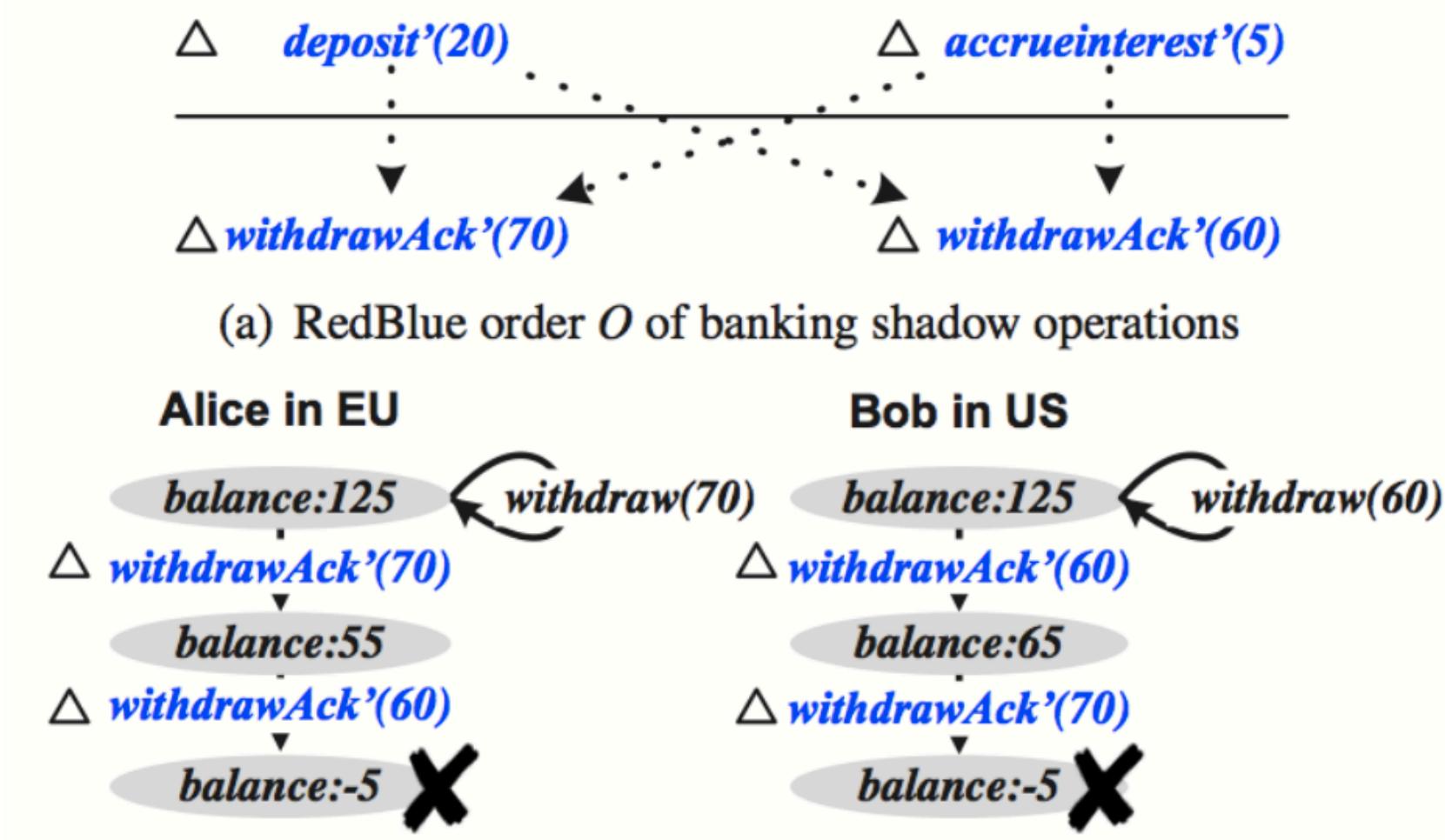
State Convergence

Convergent if all causal serializations of the RedBlue order reach the same state

All blue operations must be globally commutative

Operation Decomposition Generator & Shadow operations

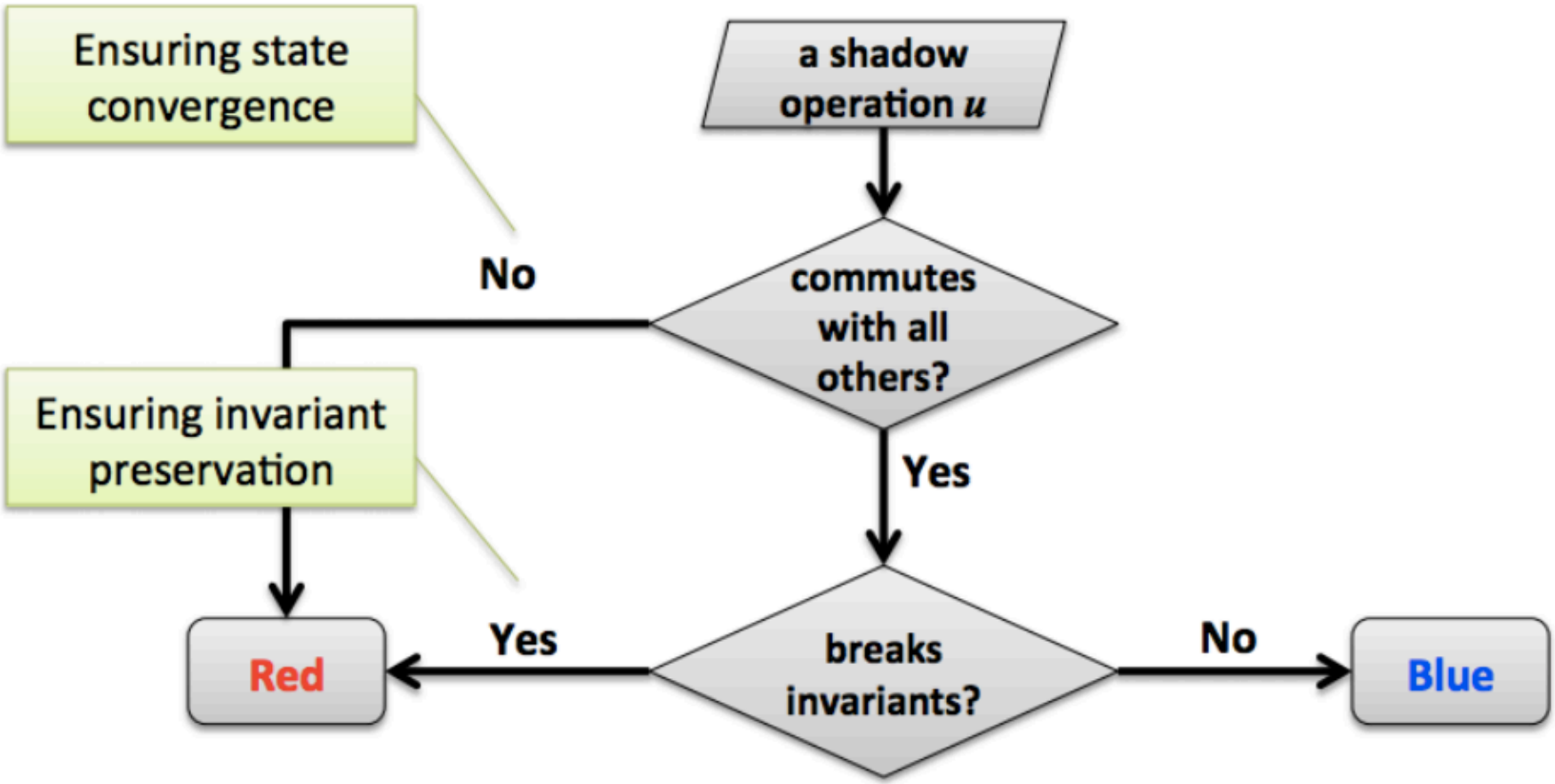
- Observation: Not all operations are commutative
- Split these operations into generator and shadow operations
- Generator Operations
  - Only executed at the primary site against a system state
  - Produces no side effects
  - Determines state transitions that would occur
  - Produces shadow operations
- Shadow Operations
  - Applies the state transitions to all the sites including the primary site
  - Must produce the same effects as the original operation given the original state for the Generator operation
- Separating operations allows for easier formation of abelian groups
- Allows for more commutative operations (blue operations)



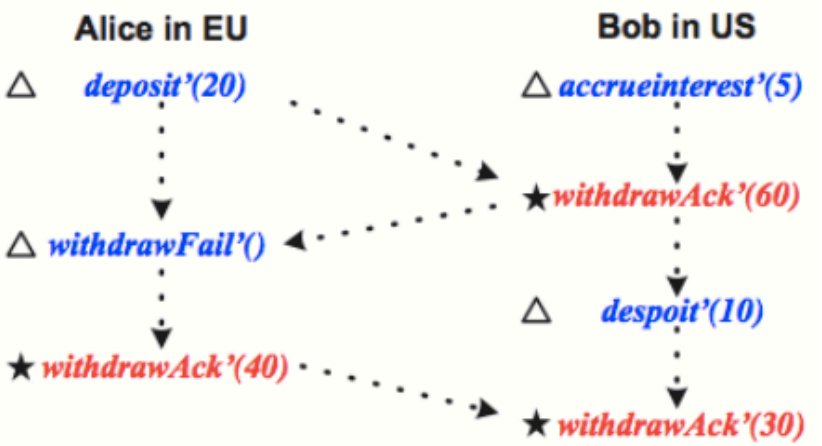
(a) RedBlue order O of banking shadow operations

Identify Red and Blue

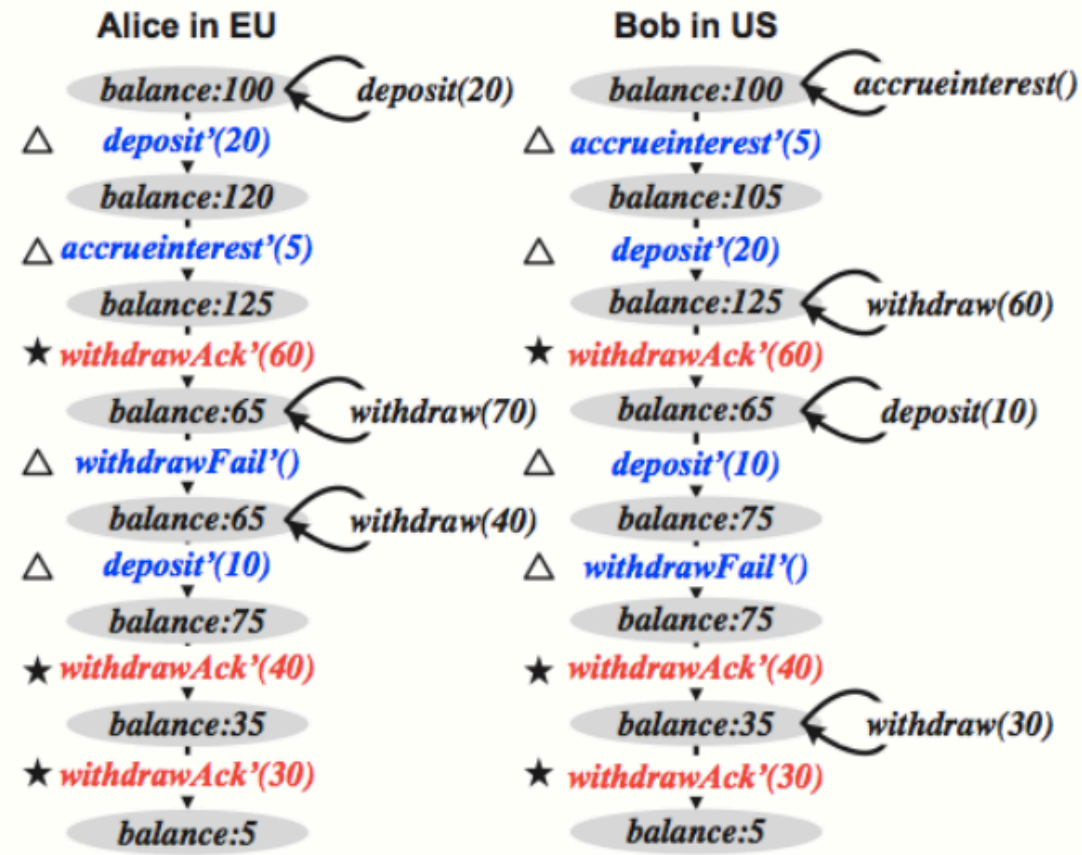
- Invariant Safe:
  - A shadow operation is invariant safe if for all valid states S and S', applying this shadow operation results in a valid state.
- If all shadow operations are correct and all blue shadow operations are **invariant safe** and **globally commutative**, then for any execution of that system that is RedBlue consistent, no site is ever in an invalid state.



Credit: author's slide



(a) RedBlue order O of banking shadow operations



(b) Convergent and invariant preserving causal serializations of O

Summary

- RedBlue consistency combines strong and eventual consistency into a single system
- The decompositon of generator/shadow operations expands the space of possible Blue operations
- A simple rule for labeling is provably state convergent and invariant preserving