Brief Announcement: Update Consistency in Partitionable Systems

Matthieu Perrin, Achour Mostefaoui, Claude Jard

Brief Announcement: Update Consistency in Partitionable Systems

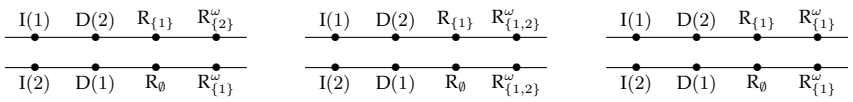
Matthieu Perrin, Achour Mostéfaoui, and Claude Jard

LINA – University of Nantes, 2 rue de la Houssinière, 44322 Nantes Cedex 3, France
matthieu.perrin@univ-nantes.fr
achour.mostefaoui@univ-nantes.fr
claude.jard@univ-nantes.fr

Data replication is essential to ensure reliability, availability and fault-tolerance of massive distributed applications over large scale systems such as the Internet. However, these systems are prone to partitioning, which by Brewer's CAP theorem [1] makes it impossible to use a strong consistency criterion like atomicity. Eventual consistency [2] guaranties that all replicas eventually converge to a common state when the participants stop updating. However, it fails to fully specify shared objects and requires additional non-intuitive and error-prone distributed specification techniques, that must take into account all possible concurrent histories of updates to specify this common state [3]. This approach, that can lead to specifications as complicated as the implementations themselves, is limited by a more serious issue. The concurrent specification of objects uses the notion of *concurrent events*. In message-passing systems, two events are concurrent if they are enforced by different processes and each process enforced its event before it received the notification message from the other process. In other words, the notion of concurrency depends on the implementation of the object, not on its specification. Consequently, the final user may not know if two events are concurrent without explicitly tracking the messages exchanged by the processes. A specification should be independent of the system on which it is implemented.

We believe that an object should be totally specified by two facets: its abstract data type, that characterizes its sequential executions, and a consistency criterion, that defines how it is supposed to behave in a distributed environment. Not only sequential specification helps repeal the problem of intention, it also allows to use the well studied and understood notions of languages and automata. This makes possible to apply all the tools developed for sequential systems, from their simple definition using structures and classes to the most advanced techniques like model checking and formal verification.

Eventual consistency (EC) imposes no constraint on the convergent state, that very few depends on the sequential specification. For example, an implementation that ignores all the updates is eventually consistent, as all replicas converge to the initial state. We propose a new consistency criterion, update consistency (UC), in which the convergent state must be obtained by a total ordering of the updates, that contains the sequential order of each pro-



(a) Not EC and not UC (b) EC but not UC (c) EC and UC

Fig. 1: Three histories for a set of integers, with different consistency criteria. An event labeled ω is repeated infinitely often.

cess. Another equivalent way to approach it is that, if the number of updates is finite, it is possible to remove a finite number of queries such that the remaining history is sequentially consistent. Unlike Fig. 1a, Fig. 1b presents an eventually consistent history, as both processes read $\{1, 2\}$ once they have converged. However, it is not update consistent: in any linearization of the updates, a deletion must appear as the last update, so this history cannot converge to state $\{1, 2\}$. State $\{1\}$ is possible because the updates can be done in the order $\mathrm{I}(2), \mathrm{D}(1), \mathrm{I}(1), \mathrm{D}(2)$, so Fig. 1c, is update consistent. As update consistency is strictly stronger than eventual consistency, an update consistent object can always be used instead of its eventually consistent counterpart.

We can prove that update consistency is universal, in the sense that every object has an update consistent implementation in a partitionable system, where any number of crashes are allowed. The principle is to build a total order on the updates on which all the participants agree, and then to rewrite the history *a posteriori* so that every replica of the object eventually reaches the state corresponding to the common sequential history. Any strategy to build the total order on the updates would work. For example, this order can be built from a timestamp made of a Lamport's clock [4] and the id of the process that performed it. The genericity of the proposed algorithm is very important because it may give a substitute to composability. Composability is an important property of consistency criteria because it allows to program in a modular way, but it is very difficult to achieve for consistency criteria. A same algorithm that pilots several objects during a same execution allows this execution to be update consistent. This universality result allows to imagine automatic compilation techniques that compose specifications instead of implementations.

References

1. Gilbert, S., Lynch, N.: Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. ACM SIGACT News **33** (2002) 51–59
2. Vogels, W.: Eventually consistent. Queue **6** (2008) 14–19
3. Burckhardt, S., Gotsman, A., Yang, H., Zawirski, M.: Replicated data types: specification, verification, optimality. In: Proceedings of the 41st annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages, ACM (2014) 271–284
4. Lamport, L.: Time, clocks, and the ordering of events in a distributed system. Communications of the ACM **21** (1978) 558–565