About State Recording in Asynchronous Computations (Abstract)

Roberto BALDONI, Jean-Michel HELARY, Michel RAYNAL
IRISA, Campus de Beaulieu, 35042 Rennes Cedex, FRANCE

The *global record* notion is of primary importance, in asynchronous parallel or distributed systems. Informally, a local record is a local state selected by a process and a global record is a set of local records, one from each process of the system. Such a global record is said to be *consistent* if it has been passed through, or if it could have been passed through, by the current computation. The purpose of this paper is to provide an answer to the following question: "given a subset of local records, can these local records belong to a consistent global record of the computation ?". Such a question has been answered in the particular context of reliable point-to-point message passing systems [2]. Here, this result is extended in two directions: first, we consider a very general computational model that encompasses shared memory, reliable point-to-point communication, multicast communication and unreliable communication systems. Second, we introduce a formal frame based on the notion of record interval and on a precedence relation defined on them. Within this general model and formal frame, a necessary and sufficient condition that answers the previous question is stated and proved. Further, a corollary of this condition suggests a strategy on how to build algorithms forcing processes to select local records in order that no previously selected local record be useless (useless local records cannot belong to any consistent global records). Due to space limitations, we only give the main notations and statement. The complete development of these ideas can be found in [1](also available by e-mail to helary@irisa.fr).

Records and intervals. A system consists of a finite set of n interacting processes $\{P_1, \ldots, P_n\}$. Each process P_i runs a program whose execution is modeled by a sequence C_i of events, called a local computation of P_i. In the following, e_i^s denotes the s-th event occurring at P_i. Events are either internal or interaction events, the latter are of two types: *get* and *put*; each *get* corresponds to one and only one *put* whereas to a *put* can correspond an arbitrary number of *get*. A *local record* is a prefix of a local computation. Local records of C_i are denoted $r_i^0, r_i^1, \ldots$ in such a way that $r_i^s \subset r_i^{s+1}$. A *global record* R of a computation is a set of local records, one for each process. Given two successive local records r_i^s and r_i^{s+1} of process P_i, the *record interval* θ_i^s is the set of events produced by process P_i between r_i^s and r_i^{s+1}.

Relations. Events of a local computation are ordered by a relation of *local precedence*, denoted $\rightarrow_l$, defined by $e_i^s \rightarrow_l e_j^t \Leftrightarrow i = j \text{ and } t > s$. Interactions between processes define on the set of events a binary relation called *get-from* and denoted $\rightarrow_g$, satisfying the property: if $e_i^s \rightarrow_g e_j^t$ then e_i^s is a *put* event and e_j^t is a *get* event. The relation of causality, denoted $\rightarrow$, is the transitive closure of the union of $\rightarrow_g$ and $\rightarrow_l$. *An asynchronous computation* is a set of local computations, one for each process, for which the relation $\rightarrow$ is a *partial order*. Relation $\rightarrow_g$ induces a *precedence* relation $\prec$ on records intervals: $\prec$ is the reflexive and transitive closure of the relation $\theta_i^x \sim \theta_j^y$ defined by $i = j$ and $y = x + 1$, or $e_i^s \rightarrow_g e_j^t$ and $e_i^s \in \theta_i^x$ and $e_j^t \in \theta_j^y$.

Orphan events and Consistency. A *get* event is called *orphan* with respect to the ordered pair of local records (r_i^s, r_j^t) if it belongs to r_j^t while the corresponding put event does not belong to r_i^s. An ordered pair (r_i^s, r_j^t) of local records is consistent iff there are no orphan *get* event with respect to this pair. A Global Record, $\{r_1^{x_1}, r_2^{x_2}, \ldots, r_n^{x_n}\}$ is *consistent* iff, for every (i, j) such that $1 \leq i \neq j \leq n$, the ordered pair of local records $(r_i^{x_i}, r_j^{x_j})$ is consistent.

Main result. The main result obtained in this paper is now stated:

Let $\mathcal{I} \subseteq \{1, \ldots, n\}$ and $\mathcal{R} = \{r_i^{x_i}\}_{i \in \mathcal{I}}$ be a set of local records of an asynchronous computation. Then $\mathcal{R}$ is a subset of a consistent global record if and only if:

$$\forall i, \forall j : i \in \mathcal{I}, j \in \mathcal{I} :: \neg(\theta_i^{x_i} \prec \theta_j^{x_j - 1})$$

Based on this result, the full paper [1] states a characterization of useless local records, shows that with each useless local record are associated particular record intervals called *pivot intervals* and describes a strategy on how to design algorithms ensuring no useless record is selected.

PODC'96, Philadelphia PA, USA

References

[1] R. Baldoni, J.M. Helary, M. Raynal, Mutually Consistent Recording in Asynchronous Computations, IRISA, Research Report no. 981, January 1996.

[2] R.H.B. Netzer, J. Xu, Necessary and sufficient conditions for consistent global snapshots, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 6,2, 1995, pp.165-169.