

# The Notion of a Timed Register and its Application to Indulgent Synchronization

Michel Raynal  
IRISA, Campus Beaulieu  
35042 Rennes, France  
raynal@irisa.fr

Gadi Taubenfeld  
Interdisciplinary Center, P.O.Box 167  
Herzliya 46150, Israel  
tgadi@idc.ac.il

## ABSTRACT

A new type of shared object, called *timed register*, is proposed and used for the design of indulgent timing-based algorithms. A timed register generalizes the notion of an atomic register as follows: if a process invokes two consecutive operations on the same timed register which are a read followed by a write, then the write operation is executed only if it is invoked at most  $d$  time units after the read operation, where  $d$  is defined as part of the read operation. In this context, a timing-based algorithm is an algorithm whose correctness relies on the existence of a bound  $\Delta$  such that any pair of consecutive constrained read and write operations issued by the same process on the same timed register are separated by at most  $\Delta$  time units. An indulgent algorithm is an algorithm that always guarantees the safety properties, and ensures the liveness property as soon as the timing assumptions are satisfied. The usefulness of this new type of shared object is demonstrated by presenting simple and elegant indulgent timing-based algorithms that solve the mutual exclusion,  $\ell$ -exclusion, adaptive renaming, test&set, and consensus problems. Interestingly, timed registers are universal objects in systems with process crashes and transient timing failures (i.e., they allow building any concurrent object with a sequential specification). The paper also suggests connections with schedulers and contention managers.

## Categories and Subject Descriptors

D.1.3 [Programming techniques]: -Concurrent Programming D.4.1 [Operating Systems] Process Management - concurrency, multiprocessing, synchronization; F.1.1 [Computation by Abstract Devices]: Models of Computation - automata, relations among models

**General Terms:** Algorithms, Reliability, Theory

**Keywords:** Asynchronous shared memory system, Atomic register, Concurrent object, Consensus, Contention man-

ager, Mutual exclusion, Process crash, Renaming, Simplicity, Test&Set, Timing assumption, Timing constraint, Universal object, Wait-free implementation.

## 1. INTRODUCTION

**Context of the study.** Concurrent programs are made up of processes that synchronize and cooperate through shared objects. These programs are executed on top of an underlying system that usually exhibits a significant degree of synchrony. This synchrony can be translated for the upper layer processes as time upper bounds in such a way that the local algorithms executed by the processes can benefit from these bounds in order to synchronize their behavior or cooperate in a correct way. Very interestingly, such synchrony assumptions allow designing algorithms with properties that could be impossible to guarantee in fully asynchronous systems.

However, program designers cannot afford to risk incorrect behavior in the periods where the synchrony assumptions are occasionally violated. An *indulgent* algorithm is an algorithm designed with synchrony assumptions in mind, that works correctly whenever the assumptions are satisfied, but never violates its safety property when these assumptions are not satisfied by the underlying system [10, 12, 13, 28]. More precisely, (1) be the assumptions it relies on satisfied or not, such an algorithm never violates its safety property, and (2) it satisfies its liveness property whenever the synchrony assumptions it relies on are satisfied.<sup>1</sup>

**Related work (1): timing-based algorithms.** Lots of synchronization algorithms have been designed that assume an upper bound  $\Delta$  on the time that can elapse between any two consecutive accesses to the shared memory issued by a process. The bound  $\Delta$  can be a priori known or not. Such algorithms are built from base atomic registers and are usually called *timing-based* algorithms. One of the very first timing-based algorithms is Fischer's mutual exclusion algorithm. This very simple and elegant algorithm (described in [20]) assumes that  $\Delta$  is known.

Fundamental results on the minimal cost (in terms of the number of atomic registers used) inherent to timing-based synchronization algorithms are stated and proved in [2, 3,

<sup>1</sup>*Safety* properties (such as, mutual exclusion, never deciding on conflicting values, if termination claimed then the application has terminated) stipulates that "nothing bad happens", while *liveness* properties (such as, eventually entering a critical section, eventually reaching an agreement, eventually claiming termination if the application terminates) stipulates that "something eventually happens" [18].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SPAA'07, June 9–11, 2007, San Diego, California, USA.

Copyright 2007 ACM 978-1-59593-667-7/07/0006 ...\$5.00.

22]. The interested reader will find in [29] a pedagogical description of several families of timing-based algorithms. Timing-based consensus algorithms are described in [2, 3, 28]. It is important to notice that the timing assumption used in these algorithms is on all the processes and all their shared memory accesses. Moreover, except for very few of them [28], these algorithms are not indulgent.

A timing-based mutual exclusion algorithm is presented in [22], in which only the deadlock-freedom property depends on the timing assumptions (mutual exclusion is guaranteed even in presence of timing failures). However (as pointed out in [28]), that algorithm is not “fully” indulgent. More precisely, the deadlock that can occur when the timing assumptions it relies on are violated, can persist even after these timing assumptions are again satisfied (i.e., even after there are no more timing failures). Furthermore, that algorithm uses only two atomic registers, and it follows from a result in [22] that any indulgent timing-based mutual exclusion algorithm for  $n$  processes, using only atomic registers must use at least  $n$  atomic registers, even when the bound  $\Delta$  is known.

**Content of the paper.** Indulgent timing-based algorithms are usually difficult to design. This comes from the fact that the safety property associated with the problem that the algorithm aims at solving has to be guaranteed not only when the timing assumptions are satisfied, but also when they are not.

The aim of the paper is to explore new ways for the design of efficient, elegant, and provably correct indulgent timing-based algorithms. To that end, a new kind of shared atomic object, called *timed register*, is first proposed, and a matching *timing assumption* is defined. Then, these notions are investigated and used to design two families of indulgent timing-based algorithms, one family is devoted to synchronization problems, while the other focuses on agreement problems.

The concept of a timed register generalizes the classical notion of an atomic register in the sense that, for each pair made up of a timed register and a process, it imposes time constraints on some write operations in order they be successful. More specifically, given a timed register  $Y$ , and a process  $p$  that invokes a read operation on  $Y$ , let us consider the first write operation on  $Y$  issued by  $p$  that follows this read (this write is called a *constrained* write, and the read is called a *constraining* read). This constrained write is executed by the timed register (it is then successful) only if it occurs at most  $d$  time units after the read, where  $d$  is a parameter defined as part of the constraining read operation. Let us notice that, not only a process may issue operations on other objects between a read and the following constrained write, but also other processes may access  $Y$ .

It is important to notice that timed registers allow restricting the timing assumptions. Classical timing-based algorithms assume a *global* timing assumption, that is, for each process, on any pair of consecutive operations, even when executed on different objects, and regardless of their type. Our timing assumption is *local*, in the sense that it applies only to a constraining read followed by a constrained write issued by the same process on the same register. We notice that, according to our timing assumption, there are no timing dependencies between operations that are invoked on different

objects nor between operations other than a read followed by a write issued by the same process on the same register. Moreover, as already indicated, a process is allowed to access other objects between a read and a write operations on the same timed register. It is worth noticing that, when all the read operations on a timed register  $Y$  specify a duration  $d = +\infty$ , all write operations on  $Y$  are successful, and the register  $Y$  boils down to a classical atomic register. This shows the way according to which the timed register notion generalizes the timed-free atomic register notion: while the latter has a fully asynchronous sequential specification, the specification of the former is also sequential but involves an additional time duration notion.

In order for a timed register  $Y$  to be useful, it must be the case that not all the write operations on  $Y$  fail. This is captured by what we call the  $\Delta$  *assumption*. This assumption states that there is a bound  $\Delta$  such that any constrained write on a timed register  $Y$  issued by a process  $p$  is executed at most  $\Delta$  time units after the immediately preceding read issued by  $p$  on  $Y$ . Consequently, the write is always successful when the duration specified by the read operation is  $\geq \Delta$ . This means that when  $\Delta$  is known and the  $\Delta$  assumption is permanently satisfied by the underlying system, any timed register behaves as an atomic register.

Unfortunately,  $\Delta$  is not always known, and in a realistic setting, the underlying system does not always guarantee the  $\Delta$  assumption (it can satisfy it only during “stable” periods). Assuming a bound  $\Delta$ , a timing failure occurs when a read of a timed register and the following constrained write are separated by more than  $\Delta$  time units. If this occurs only intermittently, the failure is *transient*. So, in our context, an indulgent timing-based algorithm has first to approximate  $\Delta$  when this bound is not known. Then, because such an algorithm preserves the safety property even in presence of transient timing failures, there is no harm in the periods where such timing failures occur. As soon as the  $\Delta$  assumption is again satisfied, the algorithm resumes its “normal” operation. Moreover, on a more engineering side, the fact that an indulgent algorithm tolerates timing failures, allows the system designer to estimate  $\Delta$  optimistically, and ignore the possible delays resulting from page faults, memory contention and preemption.

As announced previously, the paper addresses two classes of fundamental problems, namely, synchronization problems and agreement problems. More precisely, the paper first presents simple indulgent algorithms based on timed registers that solve synchronization problems, such as mutual exclusion,  $\ell$ -exclusion and wait-free adaptive renaming, in system where the bound  $\Delta$  is known. Then it presents indulgent wait-free consensus algorithms and test&set (i.e., election) algorithms for systems where the bound  $\Delta$  is known or unknown. As timed registers allow solving, in a wait-free manner, the consensus problem in systems prone to transient timing failure and where the bound  $\Delta$  is unknown, it follows that timed registers are *universal objects* in these systems, i.e., they allow building any object with an asynchronous sequential specification [14]<sup>2</sup>.

<sup>2</sup>An object  $O$  is universal if any object which has sequential specification has wait-free linearizable implementation using atomic registers and objects of type  $O$ . In [14], it is proved that consensus is universal. This implies that any object

Related work (2): contention managers and wait-free algorithms. A *wait-free implementation* of an object guarantees that any correct process (i.e., a process that does not crash) eventually terminates its operation on the object regardless of the behavior of the other processes [14]. A weaker notion that has been proposed is the notion of *obstruction-free implementation* of an object [15]. This notion stipulates that a correct process that invokes an operation on an object, and from some point in time executes alone, must eventually complete its operation. Interestingly, the weakest failure detector class that allows boosting an obstruction-free implementation into a wait-free implementation has recently been identified [11] (it is the class of eventually perfect failure detectors [7]). An interesting transformation is presented in [8] that converts any obstruction-free algorithm into a wait-free algorithm when analyzed in the unknown bound timing-based model. The notion of *abortable* objects has also been recently introduced [1]. An abortable object behaves like an ordinary object when accessed sequentially, but may return  $\perp$  when accessed concurrently.

The “undesired” behavior of an obstruction-free object implementation or an abortable object (the operation does not return or returns  $\perp$ ) can only occur in presence of concurrency. In both cases, operation liveness is guaranteed only when a single operation at a time is executed. So these notions are contention-related. They are not time-oriented.

Differently, the timed object notion as introduced in this paper is “dual” in the sense that it is related only to the speed of a process with respect to some object, i.e., regardless of the current concurrency degree and the speed of the other processes. Moreover, when considering an indulgent timing-based algorithm, the existence of a bound  $\Delta$  that makes successful the constrained write operations on a timed register, plays the role of a contention manager that guarantees that operations always terminate.

More generally, at some abstraction level, both contention managers and the existence of a bound  $\Delta$  can be seen as schedulers that provide appropriate fairness rules in order for each operation issued by a correct process to terminate.

**Roadmap.** The paper is made up of 5 sections. Section 2 presents the computational model that consists of timed and time-free atomic registers. That section also defines the notions of a timing-based algorithm, a timing failure and an indulgent algorithm. Then, assuming the bound  $\Delta$  is known, Section 3 presents indulgent algorithms for basic synchronization problems (mutual exclusion,  $\ell$ -exclusion and adaptive renaming, respectively). Section 4, focuses on classical agreement problems (test&set and consensus) when the bound  $\Delta$  is known and when it is unknown. Finally, Section 5 provides concluding remarks.

## 2. COMPUTATIONAL MODEL

### 2.1 Process model

The system is made up of a finite number of processes denoted  $p_1, p_2, \dots$ . Each process  $p_i$  has a unique index  $i$ . When the number of processes is fixed,  $n$  denotes the total number of processes. Moreover, when solving some classes of problems (mainly the class of agreement problems), we that can implement wait-free consensus is also universal.

assume that a process may crash. A process that crashes stops its execution in a definitive manner. In addition to the usual statements, a process  $p_i$  can execute the statement  $\text{delay}(d)$  where  $d$  is a time duration. Its effect is to delay  $p_i$  for an arbitrary but finite period longer than  $d$  time units. All the durations are measured with the same time unit. (Notice that this does not require the processes to have synchronized clocks; local clocks with the same drift are sufficient.)

### 2.2 Communication model

Processes communicate by reading and writing shared registers. Shared registers are denoted with uppercase letters. Local registers are denoted with lowercase letters with the process index appearing as a subscript. There are two types of shared registers.

**Time-free register.** A time-free register is nothing else than a classical atomic register which support atomic read and atomic write operations [17, 19].

The statement  $\ell_i \leftarrow X$  denotes a read of the shared register  $X$  by  $p_i$ , and the value obtained is assigned to  $p_i$ 's local variable  $\ell_i$ . Similarly,  $X \leftarrow v$  denotes the writing of the value  $v$  into  $X$ .

**Timed register.** A timed register supports atomic read and atomic write operations. A read operation of a timed register  $Y$  is denoted  $Y.\text{read}(d)$  (where  $d$  is a time duration), and always returns the current value of  $Y$ . A write operation is denoted  $Y.\text{write}(v)$ .

In order to define the semantics of a write operation on a timed register, let us consider the sequence of read and write operations issued by a process  $p_i$  on a timed register  $Y$ . A write operation on  $Y$  issued by  $p_i$  is *constrained* if it immediately follows a read operation by  $p_i$  on  $Y$  in that sequence (that read operation is the *associated constraining read*).

Let  $\text{wr}_i = Y.\text{write}(v)$  be a constrained write, and  $\text{rd}_i = Y.\text{read}(d)$  the associated constraining read. The write  $\text{wr}_i$  is *successful* if it is issued at most  $d$  time units after  $\text{rd}_i$ . In that case,  $v$  is written into  $Y$  and  $\text{wr}_i$  returns *true*. Otherwise, the write is not executed and  $\text{wr}_i$  returns *false*. Timed registers can be implemented from time-free registers and timers, but the important point here is that (as failure detector classes, or abstract data types), a timed register is defined only by behavioral properties and not by the way these properties are implemented in a particular system<sup>3</sup>.

It follows from the previous definition, that a read operation such that  $\text{rd}_i = Y.\text{read}(+\infty)$  imposes no constraint on the next write on  $Y$  issued by the same process. More generally, as already observed, if all the read operations on a timed register are such that  $d = +\infty$ , that register behaves as an atomic time-free register.

It is important to notice that, between the  $\text{rd}_i$  and  $\text{wr}_i$  operations as defined above, (1)  $p_i$  can issue read and write operations on any (time-free or timed) register different from

<sup>3</sup>Instead of constraining only the first write that appears after a read issued by the same process, one could envisage to constrain all the write operations issued by the same process and that appear between two reads of the same register. As this requirement is stronger than necessary for solving the problems we are interested in, we do not consider here this stronger definition.

$Y$ , and (2) all the other processes can issue operations on any register (including  $Y$ ).

**Remark.** A *sticky* bit [24] is a register initialized to a default value  $\perp$  that can then contain either 0 or 1. If several processes are concurrently trying to write into a sticky bit, only one of them succeeds. A write operation returns *failure* if the value it is trying to write disagrees with the already written value, and *success* otherwise. Both timed registers and sticky bits are atomic objects that provide the processes with a write operation that returns a boolean value. But, they are objects of different types as, while a sticky bit is not time-constrained but allows the write of a single value only to win, a timed register is time-constrained but does not prevent different values from being successfully written.

**Read/write pattern associated with a timed register.** Let  $Y$  be a timed register initialized to a default value  $\perp$ , and  $\delta$  a time duration. The following read/write pattern appears in a lot of algorithms based on timed registers: “**if** ( $Y.\text{read}(\delta) = \perp$ ) **then**  $Y.\text{write}(v)$  **end if**”<sup>4</sup>. This simple pattern is very often embedded in the following more general pattern made up of a small set of statements. This pattern (where  $v_i$  is the value that  $p_i$  intends to write into  $Y$ ) can be executed concurrently by several processes (it implicitly assumes that  $Y$  is not used elsewhere). As we will see, in one way or another, all our algorithms are based on this pattern:

**while** ( $Y.\text{read}(\delta) = \perp$ ) **do**  $Y.\text{write}(v_i)$  **end while**;  $\text{delay}(\delta)$ .

Interestingly, this pattern has the following property. Due to page limitation, its proof is omitted. It can be found in [27].

**THEOREM 1.** *Let us assume that the timed register  $Y$  is not accessed outside the pattern, the pattern being executed by any number of processes, and  $v_i \neq \perp$  for all  $i$ . (1) If a process reads  $Y$  after the repeat statement, it always obtain a non- $\perp$  value. (2) All the processes that read  $Y$  after the delay statement obtain the same non- $\perp$  value.*

### 2.3 Timing assumption, timing failure and indulgent timing-based algorithm

**Timing assumption.** A system satisfies a timing assumption if there is a bound  $\Delta$  such that, for each timed register  $Y$  and each process  $p_i$ , the time that separates a  $\text{rd}_i = Y.\text{read}(d)$  operation and the associated constrained write  $\text{wr}_i = Y.\text{write}(v)$  (if any) are separated by at most  $\Delta$  time units. That bound can be known or not known. If it is known,  $p_i$  can use it in the read operation, namely it can always invoke  $\text{rd}_i = Y.\text{read}(\Delta)$ . Otherwise,  $p_i$  has to approximate it by modifying the value of  $d$  until it becomes  $\geq \Delta$ . When  $d \geq \Delta$ , all the constrained write operations become successful in the absence of timing failures.

In our context, a timing-based algorithm is an algorithm whose liveness property relies on the existence of a such a bound  $\Delta$ . Nearly all the timing-based algorithms encoun-

<sup>4</sup>This pattern shows that a pair made up of a constraining read and the corresponding constrained write operation on a timed register, can be seen as a `compare&swap()` operation (1) the atomicity of which has been broken by “cutting” the `compare&swap()` operation in two operations, and (2) where the second operation is executed only if timing constraints are satisfied.

tered in the literature assume that the bound  $\Delta$  holds permanently from the very beginning and is on any pair of consecutive operations issued by a process. Differently, in this paper, no timing assumption is assumed when a process invokes operations on time-free objects, on different timed objects, or when a write follows another write on the same timed object.

**Timing failure and indulgent timing-based algorithm.** A timing failure refers to the situation where the timing assumption the underlying system should provide the processes with, is not satisfied. Here, a timing failure means that there are periods during which there is no bound  $\Delta$  such that at most  $\Delta$  time units separate any constrained write from its associated constraining read operation.

A timing-based algorithm is indulgent with respect to transient timing failures, if it never violates the safety property of the problem it has to solve even in the presence of timing failures, and satisfies its liveness property if, after some time, there are no more timing failures.

**Remark.** In systems which allow processes to be preempted (as almost all modern computing systems do),  $\Delta$  must include the time spent preempted between two statements. Furthermore, in determining the value of  $\Delta$ , possible cache misses, page faults, and delays due to memory contention should be taken into account. Thus, the value of  $\Delta$  in most cases must be very large. However, because indulgent algorithms ensure safety even when timing assumptions are violated,  $\Delta$  can be estimated optimistically, ignoring possible delays resulting from page faults, memory contention and preemption. The exact value of  $\Delta$ , when evaluated optimistically, should be tuned for each individual machine architecture. It can also be changed over time if the algorithm fails to make sufficient progress.

## 3. INDULGENT SYNCHRONIZATION

This section focuses on the use of timed registers for the design of indulgent timing-based synchronization algorithms. A simple mutual exclusion is first presented. This mutual exclusion algorithm is then used as a building block to design an  $(\ell - 1)$ -resilient indulgent solution to the  $\ell$ -exclusion problem. The  $\ell$ -exclusion algorithm is in turn used as a building block to design a wait-free solution to the adaptive renaming problem, that is optimal with respect to the new name space.

### 3.1 An indulgent mutual exclusion algorithm with a known bound

**The algorithm.** The *mutual exclusion* problem is well-known. We consider here its version where (1) no two processes can be in their critical sections at the same time (safety property), and (2) if one or more processes want to enter their critical sections, at least one of them must succeed (deadlock freedom property).

An indulgent mutual exclusion algorithm is described in Figure 1. It uses a single timed register. This algorithm is actually a simple adaptation of Fischer’s non-indulgent algorithm (as described in [20]) in which the time-free atomic register is replaced by a timed register (denoted  $Y$ , and initialized to  $\perp$ ). The statement “**await condition**” is used as an abbreviation for “**while**  $\neg\text{condition}$  **do skip**”. A process  $p_i$  first waits until  $Y = \perp$  ( $Y \neq \perp$  mean that at least one

process is already competing for the critical section). It then assigns its index  $i$  to  $Y$ . In case of a timing failure, the constrained write fails, and  $p_i$  starts all over again. Otherwise the write operation is successful. In that case,  $p_i$  delays itself for at least  $\Delta$  time units. If  $Y = i$  after that period,  $p_i$  enters its critical section. Otherwise,  $p_i$  repeats the procedure. Releasing the critical section is done by resetting  $Y$  to  $\perp$ . Let us observe that only  $p_i$  can set  $Y$  equal to  $i$ .

```

operation enter_mutex( $i$ ):
(1) repeat
(2)   await ( $Y.read(\Delta) = \perp$ );
(3)   if  $Y.write(i)$  then delay( $\Delta$ ) end if
(4) until ( $Y.read(\infty) = i$ ) end repeat

operation exit_mutex():
(5)  $Y.write(\perp)$ 

```

**Figure 1: An indulgent mutual exclusion algorithm with a known bound (code for  $p_i$ )**

**THEOREM 2.** *The algorithm described in Figure 1 always satisfies the mutual exclusion safety property. It satisfies the deadlock freedom property as soon as there are no timing failures.*

Due to page limitation, the proof is omitted. It can be found in [27].

**Additional properties of the algorithm.** The previous indulgent mutual exclusion algorithm enjoys the following additional properties (where the index  $i$  of  $p_i$  is considered as its identifier).

- It works for any number of processes (assuming the processes have unique identifiers).
- It is symmetric. The only way for distinguishing processes is by comparing their identifiers. Each identifier is unique. The identifiers can only be written, read and compared for equality (they are not used to index shared registers.)
- It follows from a result in [22] that any indulgent timing-based mutual exclusion algorithm for  $n$  processes, using only time-free atomic registers must use at least  $n$  atomic registers, even when the bound  $\Delta$  is known. An indulgent timing-based mutual exclusion algorithm was presented in [28], which uses  $O(n)$  atomic registers where  $n$  is the number of processes.

The indulgent mutual exclusion algorithm presented in Figure 1 uses only one timed register. It is consequently optimal with respect to the number of timed registers. This shows that the timed register notion and the associated timing assumption can help circumventing the lower bound on the number of registers required for solving the mutual exclusion problem.

- As for the case where the bound is *not* known, it can easily be shown, by modifying the lower bound proof from [5], that any indulgent timing-based mutual exclusion algorithm for  $n$  processes using timed registers and atomic registers must use at least  $n$  such registers. There exist many asynchronous mutual exclusion algorithms using atomic registers only that match this  $\Omega(n)$  space bound [25, 29], thus using timed registers

does not seem to help in the case where the bound  $\Delta$  is unknown.

**Implementing semaphores.** It is easy to implement an *unfair semaphore* from a timed register. In an unfair semaphore, when several processes are blocked due the **down** operation they have issued on such a semaphore, there is no assumption on which process will be allowed to proceed when another process will issue an **up** operation on that semaphore [29]. The semaphore **down** operation is simply an **enter\_mutex()** invocation, while **up** is an **exit\_mutex()** invocation. As observed in [29], it is possible to solve the classical variants of the readers-writers problem (with priority either to readers or writers) as soon as one is provided with unfair semaphores.

### 3.2 An indulgent $\ell$ -exclusion algorithm with a known bound

**The problem.** The  $\ell$ -exclusion problem is a natural generalization of the mutual exclusion problem: it consists in designing an algorithm that guarantees that up to  $\ell$  processes and no more may simultaneously access identical copies of the same non-sharable resource (that is,  $\ell$  processes are permitted to be simultaneously in their critical section; the 1-exclusion problem is exactly the mutual exclusion problem). A solution is required to withstand the slow-down or even the crash (fail by stopping) of up to  $\ell - 1$  processes (i.e., it has to be  $(\ell - 1)$ -resilient).

A good example is that of a bank where people are waiting for a teller. Here the processes are the people, the resources are the tellers, and the parameter  $\ell$  is to the number of tellers. We notice that the usual bank solution, where people line up in a single queue, and the person at the head of the queue goes to any free teller, does *not* solve the  $\ell$ -exclusion problem. If  $\ell \geq 2$  tellers are free, a proper solution should enable the first  $\ell$  people in line to move simultaneously to a teller. However, the bank solution, requires them to move past the head of the queue one at a time. Moreover, if the person at the front of the line “fails”, then the people behind this person wait forever.

An  $\ell$ -exclusion algorithm must guarantees that at most  $\ell$  processes are simultaneously in their critical sections (safety property), and, when strictly less than  $\ell$  processes fail (or are delayed forever), if processes (at least one) are trying to enter their critical sections, then at least one process will enter it (deadlock-freedom property).

**The algorithm.** Solving the  $\ell$ -exclusion problem using atomic registers only requires, as in the case of mutual exclusion, at least  $n$  such registers. In the following, we present a simple  $\ell$ -exclusion algorithm using only  $\ell$  timed-registers. The algorithm (described in Figure 2) makes use of  $\ell$  (slightly modified) copies of the mutual exclusion algorithm from the previous section (Figure 1). Those are indexed from 0 to  $\ell - 1$ . The  $c$ -th copy uses a single timed register denoted  $Y[c]$ . In its entry code, a process (say  $p_i$ ) starts participating in one of these copies (say  $c_i$ ). Process  $p_i$  is allowed to enter its critical section when it wins in one of the  $\ell$  copies.

More specifically, to enter its critical section, a process  $p_i$  executes the following statements where  $c_i$  is a local variable (the initial value of which is arbitrary):

- $p_i$  first looks for a free “slot”  $c_i$ , i.e., a copy  $Y[c_i]$  such

```

operation enter_ℓ-exclusion( $i$ ):
(1) repeat
(2)   while ( $Y[c_i].\text{read}(\Delta) \neq \perp$ )
      do  $c_i \leftarrow c_i + 1 \pmod{\ell}$  end while;
(3)   if  $Y[c_i].\text{write}(i)$  then  $\text{delay}(\Delta)$  end if
(4)   until ( $Y[c_i].\text{read}(\infty) = i$ ) end repeat

operation exit_ℓ-exclusion( $c_i$ ):
(5)  $Y[c_i].\text{write}(\perp)$ 

```

**Figure 2: An indulgent  $\ell$ -exclusion algorithm with a known bound (code for  $p_i$ )**

that  $Y[c_i] = \perp$ . This means that  $p_i$  moves to the next copy  $c_i + 1 \pmod{\ell}$  if  $Y[c_i] \neq 0$ . (Differently from the mutual exclusion algorithm,  $p_i$  does not busy-wait on a single copy but on all the copies.)

- As in the base mutual exclusion algorithm,  $p_i$  assigns its index  $i$  to  $Y[c_i]$  as soon as it finds a copy such that  $Y[c_i] = \perp$ . In case of a timing failure, since  $Y[c_i]$  is a timed register, the write operation can fail. If this is the case,  $p_i$  restarts from the beginning.
- Otherwise the write operation succeeds, and, accordingly,  $p_i$  delays itself for a duration  $d$ ,  $d \geq \Delta$ . If after this delay period,  $Y[c_i]$  has not been changed (i.e.,  $Y[c_i] = i$ ),  $p_i$  is attributed the copy  $c_i$ , otherwise it moves to the next copy.

As before, releasing the critical section is done by setting  $Y[c_i]$  to  $\perp$ . The reader can notice that, as the base mutual exclusion algorithm, this algorithm works for any number of processes.

The proof that the algorithm satisfies the safety and deadlock freedom properties, are similar to the proofs of the mutual exclusion algorithm. They are left as an exercise.

### 3.3 An indulgent adaptive renaming algorithm with a known bound

The problem. For this problem, we assume that there are  $n$  processes. Moreover, an arbitrary number of these  $n$  processes can crash.

In the *long-lived  $M$ -renaming* problem, each process  $p_i$  has an initial name  $id_i$ , that is a value from a large space  $[1..N]$  (i.e., such that  $n \ll N$ ). No two processes have the same initial name. A process  $p_i$  can invoke two operations  $\text{get\_name}(id_i)$  and  $\text{release\_name}()$ . The former operation provides  $p_i$  with a new name  $new_i$ , while the latter allows  $p_i$  to release the new name it has previously acquired. At any time, any two processes that have invoked  $\text{get\_name}(id_i)$  (and not yet invoked  $\text{release\_name}()$ ) have distinct new names (safety property). Moreover, let  $p$  be the number of processes that have currently invoked  $\text{get\_name}(id_i)$  (and not yet invoked  $\text{release\_name}()$ ). These are the processes that (1) have acquired a new name and not yet released it, plus (2) the processes that are currently competing to acquire a new name. The current new name space must be  $[1..M]$  where  $M$  is a function of  $p$  (adaptivity property). Ideally, this means that, when  $x$  new names have been attributed and not yet released and a single process  $p_i$  wants to acquire a new name, it has to obtain the name  $x + 1$ , whatever its index  $i$ .

A solution to the renaming problem is required to be *wait-free*. This means that it should guarantee that every partic-

ipating process (that does not crash) will always be able to get a new name in a finite number of steps regardless of the behavior of other processes [14]. That is, the solution has to tolerate any number of process crash failures. It is proved in [6, 16] that, when one can use only time-free atomic registers,  $2p - 1$  is a lower bound on the new name space for wait-free solving the renaming problem, where  $p$  ( $p \leq n$ ) is the number of processes that want concurrently acquire new names.

The  $M$ -renaming problem with  $M = p$  is consequently unsolvable when one can use atomic registers only. This section shows that the  $p$ -renaming problem can be solved with timed registers. Said differently, timed registers not only allow circumventing the  $2p - 1$  lower bound of the adaptive renaming problem, but allow designing an optimal solution as far as the size of the new name space is concerned.

The algorithm. The adaptive algorithm based on timed registers is described in Figure 3. A new name can be seen as a resource that a process can acquire and then release. It follows from this simple observation that an indulgent wait-free adaptive renaming algorithm can be designed by adapting the previous  $\ell$ -exclusion indulgent algorithm. The only difference is the number of resources: here there are  $n$  possible new names, so, instead of an array with  $\ell < n$  entries, the algorithm uses an array of  $n$  timed registers  $Y[1..n]$ , such the register  $Y[c]$  controls the assignment of the new name  $c$ .

```

operation get_name( $id_i$ ):
(1)  $c_i \leftarrow 1$ ;
(2) repeat
(3)   while ( $Y[c_i].\text{read}(\Delta) \neq \perp$ ) do  $c \leftarrow c_i + 1$  end while;
(4)   if  $Y[c_i].\text{write}(id_i)$  then  $\text{delay}(\Delta)$  end if
(5)   until ( $Y[c_i].\text{read}(\infty) = id_i$ ) end repeat;
(6) return ( $c_i$ )

operation release_name( $c_i$ ):
(7)  $Y[c_i].\text{write}(\perp)$ 

```

**Figure 3: An indulgent wait-free adaptive renaming algorithm with a known bound (code for  $p_i$ )**

As  $n$  new names are possible, and there are  $n$  processes, whatever its initial name, a process invoking  $\text{get\_name}(id_i)$  cannot be blocked by other processes when it wants to acquire a new name. It follows that the algorithm is trivially wait-free. The number of steps a process has to execute to obtain a new name cannot be bounded, as it depends on the occurrence of timing failures. When (1) there are no timing failures, (2) no new name is attributed, and (3)  $p$  processes concurrently want to acquire a new name, a process has to executes at most  $p$  times the **repeat** loop (lines 2-5).

The algorithm adaptivity is an immediate consequence of the fact that each process starts always scanning the new name space from 1. More generally, the proof that the algorithm is correct follows from the  $\ell$ -exclusion algorithm instantiated with  $\ell = n$ .

## 4. INDULGENT AGREEMENT

This section explores the use of timed registers for the design of wait-free indulgent test&set and consensus algorithms. In both cases, algorithms are given for the case where the bound  $\Delta$  is known and the case where  $\Delta$  is unknown.

## 4.1 Indulgent test&set algorithms

The problem. Test&set bits are well known atomic objects that are usually implemented in hardware by many computer architectures. In this section we show how to implement test&set bits from timed registers. It is important to notice that the other direction is not possible. Namely, there are problems that are solvable using timed registers which are not solvable using only test&set bits and time-free atomic registers. One such problem, the wait-free consensus problem (which is covered in the next subsections) can be solved only for two processes using test&set bits and time-free atomic registers, while using timed registers enables to solve it for any number of processes.

A *test&set* object  $TS$  is a bit with initial value 1, that supports two atomic operations. The reset operation (denoted  $TS.reset()$ ) is equivalent to writing 1 into  $TS$ . The test&set operation (denoted  $TS.test\&set()$ ) atomically returns the current value of  $TS$  and then sets it to 0.

It is easy to see that, until  $TS.reset()$  is invoked, the first invocation of  $TS.test\&set()$  returns 1, while the following invocations return 0. The process that obtains the value 1 is usually called the *winner*. Let us notice that the familiar *leader election* problem can be easily solved with a test&set object  $Y$ , the winner process is the elected leader.

An indulgent test&set algorithm with a known bound. The algorithm described in Figure 4 uses a single timed register named  $Y$  (initialized to  $\perp$ ) in order to wait-free implement a test&set object  $TS$ .

To implement  $TS.test\&set()$ , a process  $p_i$  first tests  $Y$ . If  $Y = \perp$ , it assigns its index  $i$  to  $Y$ . As a result of a timing failure, that write operation may fail. If there is no timing failure, the write operation succeeds and  $p_i$  delays itself. This procedure repeats until  $Y \neq \perp$ . At this point, observing that the algorithm implementing  $TS.test\&set()$  follows the basic read/write pattern exhibited in Section 2.2, it follows from Theorem 1, that the value of  $Y$  does not change until it is reset. Once  $Y \neq \perp$ , the process  $p_i$  such that  $Y = i$  becomes the winner, while the other processes lose. The reset operation is implemented as before by resetting  $Y$  to its initial value  $\perp$ .

```

operation  $TS.test\&set()$ :
(1) while ( $Y.read(\Delta) = \perp$ ) do
(2)   if  $Y.write(i)$  then  $delay(\Delta)$  end if
(3) end while;
(4) if  $Y.read(\infty) = i$  then  $return(1)$  else  $return(0)$  end if

operation  $TS.reset()$ :
(5)  $Y.write(\perp)$ 

```

Figure 4: An indulgent implementation of a test&set bit with a known bound (code for  $p_i$ )

The most common correctness requirement used for implementing shared objects is *linearizability*, which means that although operations of concurrent processes may overlap, each operation should appear to take effect instantaneously at some time between its first and last event [17]. In the implementation of Figure 4, the linearization point of each test&set operation, for all the processes except the winner, is exactly when the test  $Y.read(\Delta) = \perp$  at line 1 fails; the linearization point of the test&set operation of the winner is when it starts.

An indulgent test&set algorithm with an unknown bound. This section presents our first algorithm where the bound  $\Delta$  is not known. The technique that is used is both simple and general enough to be employed again in the following sections to design wait-free consensus algorithms with an unknown bound. This technique, which is very classical, consists in computing successive approximations of the unknown bound.

```

operation  $TS.test\&set()$ :
(1) while ( $Y.read(d_i) = \perp$ ) do
(2)   if  $Y.write(i)$  then  $delay(\max(\{DELAY[k]_{1 \leq k \leq n}\}))$ 
(3)     else  $d_i \leftarrow d_i + 1$ ;  $DELAY[i] \leftarrow d_i$ 
(4)   end if
(5) end while;
    % Line 6 is for efficiency only (could be suppressed) %
(6)  $DELAY[i] \leftarrow 1$ ;  $d_i \leftarrow \lceil \frac{d_i}{2} \rceil$ ;
(7) if  $Y.read(\infty) = i$  then  $return(1)$  else  $return(0)$  end if

operation  $TS.reset()$ :
(8)  $Y.write(\perp)$ 

```

Figure 5: An indulgent implementation of a test&set bit with an unknown bound (code for  $p_i$ )

The algorithm is described in Figure 5. The shared array  $DELAY[1..n]$  is an array of atomic time-free registers such that  $DELAY[i]$  contains  $p_i$ 's current estimate of the unknown bound;  $d_i$  is a local variable of process  $p_i$  initialized to 1. When we look at the algorithm of Figure 4, we observe that  $\Delta$  is used only at line 2, so that line has to be appropriately modified to cope with the fact that  $\Delta$  is not known. Now, a write operation may fail not only because there is a timing failure, but also because the current estimate is smaller than  $\Delta$ . If the write operation fails the process increases its estimate for  $\Delta$  and tries again. If it succeeds,  $p_i$  delays itself for a duration equal to the maximum estimate of  $\Delta$  taken over the estimates of all the processes, i.e.,  $\max(\{DELAY[k]_{1 \leq k \leq n}\})$ . So, line 2 of Figure 4 is replaced by the lines 2-4 in Figure 5.

We notice that, as a result of timing failures,  $p_i$ 's estimate of  $\Delta$  may be much higher than its actual value. Thus, in line 6 of the algorithm,  $d_i$  is reset to  $\lceil \frac{d_i}{2} \rceil$ , so that  $p_i$  can better estimate  $\Delta$  next time. Also, in line 6,  $DELAY[i]$  is set to 1 so that in the future it will not slow down other processes when performing the delay statement in line 2. The proof of this algorithm is similar to the proof of the ‘‘corresponding’’ consensus algorithm presented in Section 4.3. So, only that second proof is given.

## 4.2 Indulgent multi-valued consensus algorithms with a known bound

The problem. The *consensus* problem consists in designing an algorithm in which all the correct processes (i.e., the processes that do not crash) reach a common decision based on their initial opinions. Each process  $p_i$  proposes a value  $v_i$  (its opinion), and the processes have to decide in such a way that (1) each correct process eventually decides (termination property), (2) no two processes decide on different values (agreement property)<sup>5</sup>, and (3) the decided value is

<sup>5</sup>This property is sometimes called *uniform* agreement, as it prevents a process that crashes from deciding differently from the correct processes.

a proposed value (validity property).

The consensus is *binary* when the set of values that can be proposed contains only two values, otherwise it is *multi-valued*. It is well-known that the consensus problem cannot be solved in asynchronous systems made up of atomic registers only, when assuming that processes may crash [9, 14, 21]. This impossibility can be circumvented by using stronger synchronization operations (like `compare&swap()`) [14], unreliable failure detectors [7, 26], restricting the input vectors that the processes can collectively propose [23], or weakening the termination property [4].

**A simple wait-free consensus algorithm.** It is interesting to note that for the known bound model, an indulgent wait-free timing-based consensus algorithm was presented in [28], which uses infinitely many time-free atomic registers. The algorithm presented in Figure 6 is surprisingly simple and uses a single timed register  $Y$  (initialized to  $\perp$ ). A process decides when it executes the `return()` statement at line 3. As the reader can notice this algorithm is nothing else than the basic read/write pattern exhibited in section 2.2, and consequently its correctness follows directly from Theorem 1.

```

operation consensus( $v_i$ ):
(1) while ( $Y.read(\Delta) = \perp$ ) do  $Y.write(v_i)$  end while;
(2)  $delay(\Delta)$ ;
(3) return( $Y.read(\infty)$ )

```

**Figure 6: An indulgent consensus algorithm with a known bound (code for  $p_i$ )**

As it can be observed (and similarly to the previous agreement algorithms), this algorithm works for any number of processes. Moreover, differently from failure detector-based consensus algorithms, this algorithm does not use rounds. These two noteworthy features are common to all the agreement algorithms presented in the paper.

**A simple fast wait-free consensus algorithm.** The previous algorithm is not *fast* in the sense that a process has always to be delayed before deciding (see line 2). A nice property of an agreement algorithm is to allow for a fast decision in “good” circumstances. Here, those are defined as being the cases where a single value is proposed and there is no timing failure.

A fast wait-free indulgent consensus algorithm is described in Figure 7. In addition to the timed register  $Y$ , this algorithm uses an array of time-free atomic boolean register  $X[1..b]$ , where  $b$  is the number of input values that can be proposed. (To simplify the presentation we assume that the values that can be proposed are the values  $1, 2, \dots, b$ .) For each  $v$ , the flag  $X[v]$ , initialized to *false*, is set to *true* when a process proposes the value  $v$  (line 1). The array  $X[1..b]$  is then used (line 3) by a process  $p_i$  to know if a value  $v$  different from its proposal  $v_i$  has been proposed. If it is the case,  $p_i$  delays itself before deciding. Otherwise it decides immediately (line 4). It is easy to see that when a single value  $v$  is proposed, the atomic register  $X[v]$  only is set, and consequently, no process is delayed. Moreover, whatever the number of distinct values that are proposed, if there is no timing failure, a process decides after 2 or 3 accesses to the timed register  $Y$ , and  $b$  accesses to the array of time-free registers  $X[1..b]$  (i.e., 4 or 5 accesses for binary consensus).

```

operation consensus( $v_i$ ):
(1)  $X[v_i] \leftarrow true$ ;
(2) while ( $Y.read(\Delta) = \perp$ ) do  $Y.write(v_i)$  end while;
(3) if ( $\exists v : v \neq v_i \wedge X[v]$ ) then  $delay(\Delta)$  end if;
(4) return( $Y.read(\infty)$ )

```

**Figure 7: An indulgent fast consensus algorithm with a known bound (code for  $p_i$ )**

**THEOREM 3.** *The algorithm described in Figure 7 is a wait-free indulgent consensus algorithm.*

**Proof** The proof of this algorithm is verbatim the proof of Theorem 4 after having replaced the  $\delta_\alpha$  durations by the constant  $\Delta$ .  $\square_{Theorem 3}$

### 4.3 Indulgent fast consensus algorithms with an unknown bound

This section focuses on the design of fast consensus algorithms built from timed registers with an unknown bound. Such an algorithm is described in Figure 8. It is basically the same as the previous algorithm (where the bound  $\Delta$  is known), with the addition of a local variable  $d_i$  (initialized to 1), and a shared variable  $DELAY[i]$  per process  $p_i$ . These variables allow the processes to compute approximations of the unknown bound  $\Delta$ .

```

operation consensus( $v_i$ ):
(1)  $X[v_i] \leftarrow true$ ;
(2) while ( $Y.read(d_i) = \perp$ ) do
(3)   if  $\neg(Y.write(v_i))$ 
       then  $d_i \leftarrow d_i + 1$ ;  $DELAY[i] \leftarrow d_i$  end if
(4) end while;
(5) if ( $\exists v : v \neq v_i \wedge X[v]$ )
       then  $delay(\max(\{DELAY[k]_{1 \leq k \leq n}\}))$  end if;
(6) return( $Y.read(\infty)$ )

```

**Figure 8: An indulgent fast consensus algorithm with an unknown bound (code for  $p_i$ )**

Due to the presence of the array  $DELAY[1..n]$ , this algorithm requires the knowledge of  $n$ . This array can be replaced by a single atomic variable (denoted  $DELAY$  and initialized to 1) where this variable is provided with an atomic `incr()` operation (i.e.,  $DELAY.incr()$  atomically increases  $DELAY$  by 1). (It is important to recall that such an atomic operation does not enable solving consensus for an arbitrary number of processes [14].) Using such an operation, we obtain the fast consensus algorithm with an unknown bound described in Figure 9. Interestingly, this modified algorithm works for any number of processes. (The reader can notice that the structure of this algorithm is the same as the one of the fast consensus algorithm described in Figure 7, where the bound  $\Delta$  is known.)

It is easy to see that the algorithm is fast: regardless of timing failures when a single value is proposed, no `delay()` statement is executed. Moreover, as soon as  $DELAY \geq \Delta$ , with no timing failures, a process decides after  $b$  accesses to the array  $X[1..b]$ , 2 or 3 accesses to the timed register  $Y$ , and 2 accesses to the atomic register  $DELAY$ .

**THEOREM 4.** *The algorithm described in Figure 9 is a wait-free indulgent consensus algorithm.*



```

operation consensus( $v_i$ ):
(1)  $X[v_i] \leftarrow true$ ;
(2) while ( $Y.read(DELAY) = \perp$ ) do
(3)   if  $\neg(Y.write(v_i))$  then  $DELAY.incr()$  end if
(4) end while;
(5) if ( $\exists v : v \neq v_i \wedge X[v]$ ) then  $delay(DELAY)$  end if;
(6) return( $Y.read(\infty)$ )

```

**Figure 9: An indulgent consensus with an unknown bound, for any number of processes (code for  $p_i$ )**

**Proof** Validity property. That property follows directly from the fact that a decided value is a value that has been written into  $Y$  (line 6), and only proposed values can be written into  $Y$  (line 3).

Wait-free (termination) property. Let us first observe, that the  $DELAY$  counter is increased until a write into  $Y$  is successful (lines 2-4). Moreover, as soon as a value is written in  $Y$ , the processes are no longer prevented from deciding. So, let us assume that no value is ever written into  $Y$ . This means that  $DELAY$  is increased and after some time  $\tau_1$  we have  $DELAY \geq \Delta$ . Let  $\tau_2$  be a time after which there are no more timing failures. Finally, let  $\tau \geq \max(\tau_1, \tau_2)$ . Clearly, after  $\tau$ , the first constrained write into  $Y$  by a process (line 2) succeeds, and from now we have forever  $Y \neq \perp$ , contradicting the initial assumption.

Agreement property. We have to show that no two processes decide different values. If no process ever writes into  $Y$ , the agreement property is trivially satisfied. So, let us assume that at least one process successfully writes into  $Y$ . Let  $W = \{p_x, p_y, \dots\}$  be the set of the processes that successfully write into  $Y$ . Moreover, let  $p_x$  be the first process that writes  $Y$ . Let us observe that, after  $p_x$  has written  $v_x$  into  $Y$ , the processes that execute line 2, find  $Y \neq \perp$  and are consequently prevented from writing into  $Y$ . So, there is a “last” value  $v$  written into  $Y$ . (As the successful write operations on  $Y$  are atomic, and we assume an arbitrary but finite number of processes, the notions of “first” and “last” are well-defined). We show that the last value written into  $Y$  is the only value that can be decided by the processes. In the following  $p_\alpha$  is any process of  $W$ , or a process of  $W$  distinct from  $p_x$  (when both  $p_x$  and  $p_\alpha$  are used in the following statements). Let  $p_y$  be any process  $\notin W$ .

- Time instant definitions.
  1. Let  $\tau_\alpha^{-1}$  be the time at which the write operation  $X[v_\alpha] \leftarrow true$  is executed by  $p_\alpha$  (line 1).
  2. Let  $\tau_\alpha^0$  be the time of the last reading of  $Y$  by  $p_\alpha$ . (As noticed before, such a read returns  $\perp$ .)  
Let  $\delta_\alpha$  be the value of  $DELAY$  used by  $p_\alpha$  in this last read statement.
  3. Let  $\tau_\alpha^1$  be the time of the successful write of  $Y$  by  $p_\alpha$ . We have  $\tau_\alpha^1 - \tau_\alpha^0 \leq \delta_\alpha$ .
  4. Let  $\tau_\alpha^2$  be the time at which  $p_\alpha$  starts reading the array  $X[1..b]$ . We have  $\tau_\alpha^{-1} < \tau_\alpha^0 < \tau_\alpha^1 < \tau_\alpha^2$ .  
Let  $\tau_y^2$  be the time at which  $p_y$  starts reading the array  $X[1..b]$ .
- As  $p_\alpha$  reads  $\perp$  from  $Y$  at  $\tau_\alpha^0$  and  $p_x$  writes it at  $\tau_x^1$ , we have  $\tau_\alpha^{-1} < \tau_\alpha^0 < \tau_x^1 < \tau_\alpha^1 < \tau_\alpha^2$ , from which we conclude that,  $\forall p_\alpha, p_\beta \in W$ :  $\tau_\alpha^{-1} < \tau_\alpha^0 < \tau_\beta^2$ .

Similarly, for a process  $p_y$  that does not write  $Y$ , we have  $Y \neq \perp$  when  $p_y$  reads  $Y$  at line 2 for the last time, and consequently we have  $\tau_x^1 < \tau_y^2$ , from which we conclude that  $\forall p_\alpha$ :  $\tau_\alpha^{-1} < \tau_\alpha^0 < \tau_y^2$ .

This means that, when any process  $p_i$  (member of  $W$  or not) starts evaluating the predicate  $\exists v : v \neq v_i \wedge X[v]$  (line 5), all the processes that write into  $Y$  have previously updated the flag of the array  $X[1..b]$  associated with the value they write.

Let us consider two cases according to the value of the predicate when evaluated by  $p_i$ .

- $\exists v : v \neq v_i \wedge X[v]$  is *false*.  
In that case a single value is ever written into  $Y$  by the processes of  $W$  (recall that the other processes do not write  $Y$ ). It follows that only that value can be decided.

- $\exists v : v \neq v_i \wedge X[v]$  is *true*.  
In that case, several values have been proposed by processes, and it is consequently possible that some of them have been written in  $Y$ .

As the predicate is satisfied, the process  $p_i$  delays itself. Moreover, as  $\forall p_\alpha \in W$ ,  $\tau_\alpha^0 < \tau_i^2$ , we conclude that the value of the duration during which  $p_i$  delays itself (line 5) is greater than  $\max\{\delta_z\}_{p_z \in S}$ . As  $\forall p_\alpha \in W$ :  $\tau_\alpha^1 \leq \tau_\alpha^0 + \delta_\alpha$ , it follows that, when this period terminates all the processes in  $W$  have written into  $Y$ . Finally, as no more write operation on  $Y$  can be issued, it follows that  $p_i$  reads and decides the last value written into  $Y$ .  $\square$ Theorem 4

## 5. CONCLUDING REMARKS

This paper has introduced and investigated the notion of a *timed register*. Such an object is an atomic register that allows imposing timing constraints on some write operations. As we have demonstrated, timed registers allow designing synchronization and agreement algorithms that are indulgent. This means that these algorithms always guarantee the safety property of the problem (they aim at solving), and guarantee their liveness property at least when the timing assumption they rest on is satisfied. To conclude we would like to emphasize the following points.

- Differently from usual timing-based algorithms, the indulgent algorithms based on timed registers are simple to design and easy to understand (and design simplicity is a first class property!).
- All the algorithms presented in the paper (be them devoted to synchronization problems or agreement problems) are very similar. They are all variants of the same basic read/write pattern. It follows that the timed register concept can be seen as a conceptual tool that allows capturing a hidden unity among these problems.
- When the bound  $\Delta$  is known and there are no timing failures, all our indulgent algorithms based on timed registers remain correct even if the timed registers are replaced with usual atomic time-free registers. This may indicate that timed registers are the right “tool” for the design of indulgent algorithms.

- Algorithms that are correct for fully asynchronous systems are trivially indulgent (as they do not require additional timing assumptions). However, such asynchronous algorithms (when they exist) are usually less space and time efficient than the corresponding timed register-based algorithms when executed in a timing-based system.
- Even when considering timing-based algorithms only, using timed registers still allow for more efficient solutions. For example, the best known timing-based indulgent wait-free consensus algorithm with a known bound using (time-free) atomic registers only uses infinitely many such registers [28]; while such an algorithm exists using a single timed register (Section 4.2).
- Timed registers are universal objects in systems that eventually satisfy timing constraints on a per timed register basis. As noticed in the introduction, this means that timed registers allows preserving safety properties, while the  $\Delta$  assumption plays the role of a contention manager that ensures the liveness property, despite concurrency. Interestingly, this approach is dual of more traditional approaches (such as obstruction-freeness) that require an operation to execute alone in order to terminate. These two approaches can be seen as two specifications of scheduling properties that allow designing wait-free algorithms for problems that are otherwise impossible to solve in asynchronous systems with time-free atomic register only and where processes can crash.

In addition to studying timed registers, it would be interesting to define and study other types of timed objects. Such timed objects can be defined in the obvious way by extending the definition of the corresponding asynchronous objects. Other ways to extend this research would be to solve other problems and design indulgent concurrent data structures using timed objects, and assume that a timed register itself can experience some kind of failures.

## 6. REFERENCES

- [1] Aguilera M.K., Frolund S., Hadzilacos V., Horn S.L. and Toueg S., Abortable Shared Objects. Brief announcement in *Proc. 20th Symp. on Distributed Computing (DISC'06)*, Springer-Verlag LNCS #4167, pp. 534-536, 2006.
- [2] Alur R., Attiya H. and Taubenfeld G., Time-adaptive Algorithms for Synchronization. *SIAM Journal of Computing*, 26(2):539-556, 1997.
- [3] Alur R. and Taubenfeld G., Fast Timing-based Algorithms. *Distributed Computing*, 10(1):1-10, 1996.
- [4] Ben-Or M., Another Advantage of Free Choice: Completely Asynchronous Agreement Protocols. *Proc. 2nd ACM Symposium on Principles of Distributed Computing (PODC'83)*, pp. 27-30, 1983.
- [5] Burns J.N. and Lynch N.A., Bounds on shared-memory for mutual exclusion. *Information and Computation*, 107(2):171-184, December 1993.
- [6] Burns J.E. and Peterson G.L., The Ambiguity of Choosing. *Proc. 8th ACM Symposium on Principles of Distributed Computing (PODC'89)*, ACM Press, pages 145-158, 1989.
- [7] Chandra T.D. and Toueg S., Unreliable Failure Detectors for Reliable Distributed Systems. *Journal of the ACM*, 43(2):225-267, 1996.
- [8] Fich E.F., Luchangco V., Moir M. and Shavit N., Obstruction-free algorithms that can be practically wait-free. *Proc. 19th Symposium on Distributed Computing (DISC'05)*, Springer-Verlag LNCS #3724, pp. 78-92, 2005.
- [9] Fischer M.J., Lynch N. and Paterson M.S., Impossibility of Distributed Consensus with One Faulty Process. *Journal of the ACM*, 32(2):374-382, 1985.
- [10] Guerraoui R., Indulgent Algorithms. *Proc. 19th ACM Symposium on Principles of Distributed Computing, (PODC'00)*, ACM Press, pp. 289-298, 2000.
- [11] Guerraoui R., Kapalka M. and Kouznetsov P., The Weakest failure Detectors to Boost Obstruction-Freedom. *Proc. 20th Symposium on Distributed Computing (DISC'06)*, Springer-Verlag LNCS #4167, pp. 376-390, 2006.
- [12] Guerraoui R. and Lynch N., A General Characterization of Indulgence. *Proc. 8th Int'l Symposium on Self-stabilization, Safety and Security of Distributed Systems (SSS'06)*, Springer-Verlag LNCS, 2006.
- [13] Guerraoui R. and Raynal M., The Information Structure of Indulgent Consensus. *IEEE Transactions on Computers*, 53(4):453-466, 2004.
- [14] Herlihy M.P., Wait-Free Synchronization. *ACM TOPLAS*, 13(1):124-149, 1991.
- [15] Herlihy M.P., Luchangco V. and Moir M., Obstruction-free Synchronization: Double-ended Queues as an Example. *Proc. 23th IEEE Int'l Conference on Distributed Computing Systems (ICDCS'03)*, pp. 522-529, 2003.
- [16] Herlihy M.P. and Shavit N., The topological Structure of Asynchronous Computability. *JACM*, 46(6):858-923, 1999.
- [17] Herlihy M.P. and Wing J.M., Linearizability: a Correctness Condition for Concurrent Objects. *ACM TOPLAS*, 12(3):463-492, 1990.
- [18] Lamport. L., Proving the Correctness of multiprocess programs. *IEEE Transaction on Software Engineering*, SE-3(2):125-143, 1977.
- [19] Lamport. L., On Interprocess Communication, Part II: Algorithms. *Distributed Computing*, 1(2):86-101, 1986.
- [20] Lamport L., A Fast Mutual Exclusion Algorithm. *ACM Transactions on Computer Systems*, 5(1):1-11, 1987.
- [21] Loui M.C. and Abu-Amara H.H., Memory Requirements for Agreement among Unreliable Asynchronous Processes. *Advances in Comp. Research*, JAI Press, 4:163-183, 1987.
- [22] Lynch N.A. and Shavit N., Timing-based Mutual Exclusion. *Proceedings of the 13th IEEE Real-Time Systems Symposium (RTSS'92)*, pp. 211, 1992.
- [23] Mostefaoui A., Rajsbaum S. and Raynal M., Conditions on Input Vectors for Consensus Solvability in Asynchronous Distributed Systems. *JACM*, 50(6):922-954, 2003.
- [24] Plotkin S.A., Sticky Bits and Universality of Consensus. *8th ACM Symposium on Principles of Distributed Computing, (PODC'89)*, ACM Press, pp. 159-175, 1989.
- [25] Raynal M., Algorithms for mutual exclusion. *The MIT Press*, ISBN 0-262-18119-3, 107 pages, 1986.
- [26] Raynal M., A Short Introduction to Failure Detectors for Asynchronous Distributed Systems. *ACM SIGACT News, Distributed Computing Column*, 36(1):53-70, 2005.
- [27] Raynal M. and Taubenfeld G., The notion of Timed Registers and its application to Indulgent Synchronization. *Tech Report RR-1820*, IRISA, Université de Rennes, France, 17 pages, October 2006.
- [28] Taubenfeld G., Computing in the Presence of Timing Failures. *Proc. 26th Int'l IEEE Conference on Distributed Computing Systems (ICDCS'06)*, 2006.
- [29] Taubenfeld G., Synchronization Algorithms and Concurrent Programming. *Pearson Prentice-Hall*, ISBN 0-131-97259-6, 423 pages, 2006.