



HAL
open science

Self-Stabilizing Wavelets and r-Hops Coordination

Christian Boulinier, Franck Petit

► **To cite this version:**

Christian Boulinier, Franck Petit. Self-Stabilizing Wavelets and r-Hops Coordination. 2007. hal-00157946

HAL Id: hal-00157946

<https://hal.science/hal-00157946>

Submitted on 27 Jun 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Self-Stabilizing Wavelets and ϱ -Hops Coordination

Christian Boulinier and Franck Petit
LaRIA, CNRS
Université de Picardie Jules Verne, France

Abstract

We introduce a simple tool called the *wavelet* (or, ϱ -wavelet) scheme. Wavelets deals with coordination among processes which are at most ϱ hops away of each other. We present a self-stabilizing solution for this scheme. Our solution requires no underlying structure and works in arbitrary anonymous networks, *i.e.*, no process identifier is required. Moreover, our solution works under any (even unfair) daemon.

Next, we use the wavelet scheme to design self-stabilizing *layer clocks*. We show that they provide an efficient device in the design of local coordination problems at distance ϱ , *i.e.*, ϱ -barrier synchronization and ϱ -local resource allocation (LRA) such as ϱ -local mutual exclusion (LME), ϱ -group mutual exclusion (GME), and ϱ -Reader/Writers. Some solutions to the ϱ -LRA problem (*e.g.*, ϱ -LME) also provide transformers to transform algorithms written assuming any ϱ -central daemon into algorithms working with any distributed daemon.

Keywords: Barrier Synchronization, Local Synchronization, Resource Allocation, Self-Stabilization, Unison.

1 Introduction

Most of the distributed system are not fully connected networks. Each process is only directly connected with a subset of others, called neighbors. By this way, the communication links organize the network in a some graph topology which is either arbitrary or in accordance with some global topology constraints, *e.g.*, acyclicity, constant degree, ring, grid, etc. Whatever the topology complexity, the design of a distributed task is simplified if it requires coordination mechanisms involving a process with its neighbors only, *i.e.*, one hop away. Such distributed tasks are said to be *local*. Unfortunately, many distributed tasks requires coordination farther away than the immediate neighbors, *i.e.*, ϱ hops away with $\varrho > 1$. If ϱ is equal to the diameter of the network D , then the task is said to be *global*.

In this paper, we consider problems requiring coordination among processes which are at most ϱ hops away of each other. We present solutions having the desirable property of self-stabilization. The concept of *self-stabilization* [Dij74, Dol00] is an efficient approach to design distributed systems to tolerate arbitrary transient faults. A self-stabilizing system, regardless of the initial states of the processors and initial messages in the links, is guaranteed to converge to the intended behavior in finite time.

Motivation and Related Works. Coordination at distance ϱ received a particular attention in recent works. There are various motivations for this issue. The first one consists in the design of

ϱ -local computations [NS93], *i.e.*, running in constant time independent of any global parameter like the size of the network or the diameter. Computation in constant time ϱ can be achieved if the processes can collect informations from processes located within radius of ϱ from them. In [NS93], the authors mainly address *Local Checkable Labeling* problems. Local computation is also considered in [GMM04] by considering the recognition problem. The computing model is a relabelling system.

Wireless networks bring new trends in distributed systems which also motivate research the local control of concurrency at distance ϱ . In [DNT06], the authors propose a generalization of the well-known dining philosophers problem [Dij68]. They extends the conflict processes beyond the immediate neighbors of the processes. As an application, their solution provide a solution to the interfering transmitter problem in wireless networks.

Another motivation consists in assuming that the knowledge of the processes goes beyond their immediate neighbors could help in the design of non-trivial tasks [GGH⁺04, GHJT06]. An efficient self-stabilizing solution is given to the *maximal 2-packing* problem assuming the knowledge at distance 2 [GGH⁺04]. (The maximal 2-packing problem consists to find a maximal set of nodes S , such that no two nodes in S are adjacent and no two nodes in S have a common neighbor.) The solution in [GGH⁺04] requires process ID's and works under a central daemon. In [GHJT06], the authors propose a ϱ -distance knowledge transformer to construct self-stabilizing algorithms which use a ϱ distance knowledge. Again, their solution works only if the daemon is central and with process ID's.

Note that various kinds of transformers have been proposed in the area of self-stabilization to refine self-stabilizing algorithms which use tight scheduling constraints like the central daemon into the corresponding self-stabilizing algorithm working assuming weaker daemons, *e.g.*, [MN98, GH99, NA02, CDP03]. A popular technique consists in composing the algorithm with a self-stabilizing *local mutual exclusion* (LME) algorithm [MN98, GH99, NA02]. LME allows to grant privileged processes to enter critical section if and only if none of their neighbors has the privilege, this infinitely often. So, Any LME-based solution does not allow concurrent execution of neighboring processes. The solution in [CDP03] is based on the *Local Resource Allocation* (LRA), which allows neighboring processes to enter their critical sections concurrently provided they do not use conflicting resources. It transforms any algorithms written in a high-atomicity model (*e.g.*, with a central daemon) into the distributed read/write atomicity model by allowing neighborhood concurrency.

However, none of the above solutions allows coordination farther than the immediate neighbors. So, they are not directly applicable to the method developed in [NS93, GGH⁺04, GHJT06, DNT06].

Contributions. In this paper, we introduce a simple tool called the *wavelet* (or, ϱ -*wavelet*) scheme. Wavelets deals with coordination among processes which are at most ϱ hops away of each other. Wavelets are related to the notion of *wave* (also called *Total* algorithm [Tel88, Tel04]).

In this paper, we present a self-stabilizing solution to the ϱ -wavelet problem. There are several way to design the wavelet scheme depending on network properties. For instance, assuming a unique identifier on each process, in [DNT06], the authors provides a self-stabilizing ϱ -wavelets scheme by combining a stabilizing Propagation of Information with Feedback (PIF) [BDPV99] over a self-stabilizing BFS spanning tree [HC92, Joh97] rooted at each process of height equal to ϱ .

By contrast, our solution requires no underlying structure and works in arbitrary anonymous networks, *i.e.*, no process identifier is required. Our solutions is based on the unison in [BPV04] and works assuming any distributed (even unfair) daemon.

Next, we use the wavelet scheme to design self-stabilizing *layer clocks*. The lower layer clock, in the sequel called the *main* clock, provides a wavelet stream. The upper layer clock, so called the *slave* clock, achieves a ϱ -*barrier* synchronization mechanism, where no process p starts to execute

its phase $i + 1$ before all processes in the ϱ -ball centered in p have completed their phase i .

Finally, we show that the layer clock also provides an efficient underlying device in the design of various local resource allocation problems at distance ϱ . This problems include Mutual Exclusion [Dij65], Group Mutual Exclusion [Jou00], and Readers-Writers [CHP71]. Some of these solutions (e.g., ϱ -LME) also provides transformers to transform algorithms written assuming any ϱ -central daemon into algorithms working with any distributed daemon.

Paper Outline. The remainder of the paper is organized as follows. We formally describe notations, definitions, and the execution model in Section 2. We also state what it means for a protocol to be self-stabilizing. In Section 3, we define the wavelet scheme, present our solution for this problem in an arbitrary anonymous networks, and show how it can be used as an infimum computation at distance ϱ . In Section 4, we introduce the self-stabilizing layer clocks and show how they can be used to solve ϱ -local coordination problems. Finally, we make some concluding remarks in Section 5.

2 Preliminaries

In this section, we first define the model of distributed systems considered in this paper. We then define the execution model and various general definitions such as events, causal DAG, and Coherent Cuts. We also state what it means for a protocol to be self-stabilizing.

2.1 Distributed System

A *distributed system* is an undirected connected graph, $G = (V, E)$, where V is a set of nodes— $|V| = n$, $n \geq 2$ —and E is the set of edges. Nodes represent *processes*, and edges represent *bidirectional communication links*. A communication link (p, q) exists iff p and q are neighbors. The distributed system is considered to be arbitrary and anonymous, *i.e.*, we consider no particular topology nor unique identifiers on processes.

The set of neighbors of every process p is denoted as \mathcal{N}_p . The *degree* of p is the number of neighbors of p , *i.e.*, equal to $|\mathcal{N}_p|$. The distance between two processes p and q , denoted by $d(p, q)$, is the length of the shortest path between p and q . Let ϱ be a positive integer. Define $V(p, \varrho)$ as the set of processes such that $d(p, q) \leq \varrho$, *i.e.*, the ϱ -ball centered at p . D denote the diameter of the network.

The program of a process consists of a set of registers (also referred to as variables) and a finite set of guarded actions of the following form: $\langle label \rangle :: \langle guard \rangle \longrightarrow \langle statement \rangle$. Each process can only write to its own registers, and read its own registers and registers owned by the neighboring processes. The guard of an action in the program of p is a boolean expression involving the registers of p and its neighbors. The statement of an action of p updates one or more registers of p . An action can be executed only if its guard evaluates to true. The actions are atomically executed, meaning the evaluation of a guard and the execution of the corresponding statement of an action, if executed, are done in one atomic step.

2.2 Execution Model

The *state* of a process is defined by the values of its registers. The *configuration* of a system is the product of the states of all processes. Let a distributed protocol \mathcal{P} be a collection of binary transition relations denoted by \mapsto , on Γ , the set of all possible configurations of the system. \mathcal{P} describes an

oriented graph $S = (\Gamma, \mapsto)$, called the *transition graph* of \mathcal{P} . A sequence $e = \gamma_0, \gamma_1, \dots, \gamma_i, \gamma_{i+1}, \dots$ is called an *execution* of \mathcal{P} iff $\forall i \geq 0, \gamma_i \mapsto \gamma_{i+1} \in S$. A process p is said to be *enabled* in a configuration γ_i ($\gamma_i \in \Gamma$) if there exists an action A such that the guard of A is true in γ_i . Similarly, an action A is said to be enabled (in γ) at p if the guard of A is true at p (in γ). We consider that any enabled processor p is *neutralized* in the computation step $\gamma_i \mapsto \gamma_{i+1}$ if p is enabled in γ_i and not enabled in γ_{i+1} , but does not execute any action between these two configurations. (The neutralization of a processor represents the following situation: At least one neighbor of p changes its state between γ_i and γ_{i+1} , and this change effectively made the guard of all actions of p false.)

We assume an *unfair and asynchronous distributed daemon*. *Unfairness* means that even if a processor p is continuously enabled, then p may never be chosen by the daemon unless p is the only enabled processor. The *asynchronous distributed daemon* implies that during a computation step, if one or more processors are enabled, then the daemon chooses at least one (possibly more) of these enabled processors to execute an action.

In order to compute the time complexity, we use the definition of *round* [DIM97]. This definition captures the execution rate of the slowest processor in any computation. Given an execution e , the *first round* of e (let us call it e') is the minimal prefix of e containing the execution of one action of the protocol or the neutralization of every enabled processor from the first configuration. Let e'' be the suffix of e , i.e., $e = e'e''$. Then *second round* of e is the first round of e'' , and so on.

2.3 Events, Causal DAG's and Cuts

Definition 2.1 (Events) Let $\gamma_0\gamma_1\dots$ be a finite or infinite execution. For all $p \in V$, $(p, 0)$ is an event. Let $\gamma_t \mapsto \gamma_{t+1}$ be a transition. If the process p executes a guarded action during this transition, we say that p executes an action at time $t + 1$. The pair $(p, t + 1)$ is said to be an event (or a p -event). Events so that the guard does not depend on the shared registers of any neighbor are said to be internal.

Definition 2.2 (Causal DAG) The causal DAG associated is the smallest relation \rightsquigarrow on the set of events such that the following two conditions hold:

1. Let (p, t) and (p, t') be two events such that $t > t_0$, t' is the greatest integer such that $t_0 \leq t' < t$. Then, $(p, t') \rightsquigarrow (p, t)$;
2. Let (p, t) and (q, t') be two events such that (p, t) is not an internal event, $q \in \mathcal{N}_p$, $t > t_0$, and t' is the greatest integer such that $t_0 \leq t' < t$. Then, $(q, t') \rightsquigarrow (p, t)$.

Denote the *causal order* on the sequence $\gamma_0\gamma_1\dots$ by \preceq . Relation \preceq is the reflexive and transitive closure of the causal relation \rightsquigarrow . The *past cone* of an event (p, t) is the causal-DAG induced by every event (q, t') such that $(q, t') \preceq (p, t)$. A past cone involves a process q iff there is a q -event in the cone. We say that a past cone *covers* V , iff every process $q \in V$ is involved in the cone. The *cover* of an event (p, t) , denoted by $Cover(p, t)$, is the set of processes q covered by the past cone of (p, t) .

Definition 2.3 (Cut) A cut C on a causal DAG is a map from V to \mathbb{N} , which associates a process p with a time t_p^C . We mix this map with its graph: $C = \{(p, t_p^C), p \in V\}$.

The *past* of C , denoted by $]\leftarrow, C]$, is the set of events (p, t) such that $t \leq t_p^C$. Similarly, we define the *future* of C , denoted by $[C, \rightarrow[$, as the set of events (p, t) such that $t_p^C \leq t$. A cut is said to be *coherent* if $(q, t') \preceq (p, t)$ and $(p, t) \preceq (p, t_p^C)$, then $(q, t') \preceq (q, t_q^C)$. A cut C_1 is less than or equal to a cut C_2 , denoted by $C_1 \preceq C_2$, if the past of C_1 is included in the past of C_2 .

If C_1 and C_2 are coherent cuts such that $C_1 \preceq C_2$, then $[C_1, C_2]$ is the *induced* causal DAG defined by the events (p, t) such that $(p, t_p^{C_1}) \preceq (p, t) \preceq (p, t_p^{C_2})$. A *sequence of events* is any segment $[C_1, C_2]$ where C_1 and C_2 are coherent cuts satisfying $C_1 \preceq C_2$. Any event of C_1 is called an *initial event*.

2.4 Self-Stabilization

Let \mathcal{X} be a set. A *predicate* P is a function that has a Boolean value—**true** or **false**—for each element $x \in \mathcal{X}$. A predicate P is *closed* for a transition graph S iff every state of an execution e that starts in a state satisfying P also satisfies P . A predicate Q is an attractor of the predicate P , denoted by $P \triangleright Q$, iff Q is closed for S and for every execution e of S , beginning by a state satisfying P , there exists a configuration of e for which Q is true. A transition graph S is *self-stabilizing* for a predicate P iff P is an attractor of the predicate **true**, *i.e.*, $\mathbf{true} \triangleright P$.

3 Wavelets

In this section, we first define the problem considered in this paper, followed by our self-stabilizing solution designed for any anonymous networks. Next, we show that it provides an efficient tool to compute any infimum at distance ϱ .

3.1 Problem Definition

Let us assume that there exists a special internal type of events called a *decide* event. Let ϱ be an integer. A ϱ -*wavelet* is a *sequence of events* $[C_1, C_2]$ that satisfies the following two requirements:

1. The causal DAG induced by $[C_1, C_2]$ contains at least one decide event;
2. For each decide event (p, t) , the past of (p, t) in $[C_1, C_2]$ covers $V(p, \varrho)$.

A *wave* is the particular case where $\varrho \geq D$, D is the diameter of the network. There are several ways to implement the ϱ -wavelet scheme if the processes have Id's, for instance using the PIF scheme on trees with height equal to ϱ rooted at each process. In the following subsection, we present a solution for the ϱ -wavelet problem in anonymous networks. Next, we show how this solution provides a self-stabilizing infimum computation in a ϱ -ball.

3.2 Solution Description

Our solution is based on the unison developed in [BPV04], which stabilizes in $O(n)$ rounds in general graphs. Note that in a tree, we could use the protocol proposed in [BPV06]. It gives the better stabilization time complexity of at most D rounds. In the sequel, we first borrow some basic definitions and properties introduced in [BPV04], followed by our solution and its correctness proof.

3.2.1 Unison

Basic Definitions and Properties. Let \mathbb{Z} be the set of integers and K be a strictly positive integer. Two integers a and b are said to be *congruent modulo* K , denoted by $a \equiv b[K]$ if and only if $\exists \lambda \in \mathbb{Z}, b = a + \lambda K$. We denote \bar{a} the unique element in $[0, K - 1]$ such that $a \equiv \bar{a}[K]$. $\min(\overline{a - b}, \overline{b - a})$ is a *distance* on the torus $[0, K - 1]$ denoted by $d_K(a, b)$. Two integers a and b are said to be *locally comparable* if and only if $d_K(a, b) \leq 1$. We then define the *local order relationship*

\leq_l as follows: $a \leq_l b \stackrel{\text{def}}{\iff} 0 \leq \overline{b-a} \leq 1$. If a and b are two locally comparable integers, we define $b \ominus a$ as follows: $b \ominus a \stackrel{\text{def}}{=} \overline{b-a}$ if $a \leq_l b$ else $-\overline{a-b}$. If $a_0, a_1, a_2, \dots, a_{p-1}, a_p$ is a sequence of integers such that $\forall i \in \{0, \dots, p-1\}$, a_i is locally comparable to a_{i+1} , then $S = \sum_{i=0}^{p-1} (a_{i+1} \ominus a_i)$ is the *local variation* of this sequence.

Define $\mathcal{X} = \{-\alpha, \dots, 0, \dots, K-1\}$, where α is a positive integer. Let φ be the function from \mathcal{X} to \mathcal{X} defined by: $\varphi(x) \stackrel{\text{def}}{=} \overline{x+1}$ if $x \geq 0$ else $x+1$. The pair (\mathcal{X}, φ) is called a *finite incrementing system*. K is called the *period* of (\mathcal{X}, φ) . Let $\text{tail}_\varphi = \{-\alpha, \dots, 0\}$ and $\text{ring}_\varphi = \{0, \dots, K-1\}$ be the sets of “extra” values and “expected” values, respectively. The set tail_φ^* is equal to $\text{tail}_\varphi \setminus \{0\}$. A *reset* on \mathcal{X} consists in enforcing any value of \mathcal{X} to $-\alpha$. We assume that each process p maintains a clock register $p.r$ using an incrementing system (\mathcal{X}, φ) . Let γ be a system configuration, we define the predicate WU as follows: $WU(\gamma) \stackrel{\text{def}}{=} \forall p \in V, \forall q \in \mathcal{N}_p : (p.r \in \text{ring}_\varphi) \wedge (|p.r - q.r| \leq 1)$ in γ . In the remainder, we will abuse notation, referring to the corresponding set of configurations simply by WU .

In WU , the clock values of neighboring processes are locally comparable. In the sequel of the paper, we need the three following definitions:

Delay. The delay of a path $\mu = p_0 p_1 \dots p_k$, denoted by δ_μ , is the local variation of the sequence $p_0.r, p_1.r, \dots, p_k.r$, *i.e.*, $\delta_\mu = \sum_{i=0}^{k-1} (p_{i+1}.r \ominus p_i.r)$ if $k > 0$, 0 otherwise ($k = 0$).

Intrinsic Delay. The delay between two processes p and q is *intrinsic* if it is independent on the choice of the path from p to q . The delay is *intrinsic* iff it is *intrinsic* for every p and q in V . In this case, and at time t , the intrinsic delay between p and q is denoted by $\delta_{(p,q)}$.

WU_0 . The predicate WU_0 is true for a system configuration γ iff γ satisfies WU and the delay is intrinsic in γ .

Unison Definition. Assume that each process p maintains a register $p.r \in \mathcal{X}$. The self-stabilizing asynchronous unison problem, or simply the *unison* problem, consists in the design of a protocol so that the following properties are true in every execution [BPV05]:

Safety : WU is closed.

Synchronization: In WU , a process can increment its clock $p.r$ only if the value of $p.r$ is lower than or equal to the clock value of all its neighbors.

No Lockout (Liveness): In WU , every process p increments its clock $p.r$ infinitely often.

Convergence: $\Gamma \triangleright WU$.

The following guarded action solves the *synchronization property* and the *safety*:

$$\forall q \in \mathcal{N}_p : (q.r = p.r) \vee (q.r = \varphi(p.r)) \longrightarrow p.r := \varphi(p.r)$$

The predicate WU_0 is closed for any execution of this guarded action. Moreover, for any execution starting from a configuration in WU_0 , the *no lockout property* is guaranteed. Generally, this property is not guaranteed in WU .

if $k \geq \perp_0 + D$, then $t_{p,k}$ is well defined and the cut $C_k = \{(p, t_{p,k}), p \in V\}$ is well defined on the network.

We now need to prove that for every $k \geq \perp_0 + D$, the cuts C_k are coherent. We first claim the following obvious lemma:

Lemma 3.1 *If $(p, t) \rightsquigarrow (q, t')$ then: $\widetilde{q^{t'}.r} \in \{\widetilde{p^{t'}.r}, \widetilde{p^{t'}.r} + 1\}$*

Inductively, if $(q_0, t_0) \rightsquigarrow (q_1, t_1) \rightsquigarrow (q_2, t_2) \dots \rightsquigarrow (q_i, t_i)$ then: $\widetilde{q_i^{t_i}.r} \in \{\widetilde{q_0^{t_0}.r}, \dots, \widetilde{q_0^{t_0}.r} + i\}$

From the Lemma 3.1, if $(q, t) \preceq (p, t_{p,k})$ then $(q, t) \preceq (q, t_{q,k})$. It follows:

Lemma 3.2 *For every $k \geq \perp_0 + D$ the cut C_k is coherent.*

Lemma 3.3 *Let $k \geq \perp_0 + D$. If (p, t) is an event in the interval $[C_k, \rightarrow[$, then $V(p, \widetilde{p^{t'}.r} - k) \subset \text{Cover}(p, t)$.*

Proof. The statement holds for the initial events of $[C_k, \rightarrow[$. Let \mathcal{A} be the set of events (p, t) in $[C_k, \rightarrow[$ such that $V(p, \widetilde{p^{t'}.r} - k) \subset \text{Cover}(p, t)$ does not hold. We assume that \mathcal{A} is not empty, let (q, τ) a minimal event in \mathcal{A} according to \preceq . Let $\varrho = \widetilde{q^{t'}.r} - k$, and let $p_1 \in V(q, \varrho)$. If $p_1 = q$ then $p_1 \in \text{Cover}(q, \tau)$, otherwise there exists $q_1 \in \mathcal{N}_q$ such that $p_1 \in V(q_1, \varrho - 1)$. (q, τ) is not an initial event, so $q_1 \in \mathcal{N}_q^\tau$ and there exists τ_{q_1} such that $(q_1, \tau_{q_1}) \rightsquigarrow (q, \tau)$. By the minimality of (q, τ) , $V(q_1, \varrho - 1) \subset \text{Cover}(q_1, \tau_{q_1})$ holds. So, $V(q_1, \varrho - 1) \subset \text{Cover}(q, \tau)$. Thus, $p_1 \in \text{Cover}(q, \tau)$. Therefore, (q, τ) is not in \mathcal{A} . Thus $\mathcal{A} = \emptyset$, and the lemma is proved. \square

As a corollary of Lemma 3.3, the following result holds :

Theorem 3.4 *Let $k \geq \perp_0 + D$ and ϱ be a positive integer. Then, $[C_k, C_{k+\varrho}]$, with $C_{k+\varrho}$ as the set of decide events, is a ϱ -wavelet.*

3.3 Infimum Computation

Problem definition. In [Tel88, Tel04], the author introduces the infimum operators. An infimum \oplus over a set \mathbb{S} , is an associative, commutative and idempotent (i.e. $x \oplus x = x$) binary operator. If $P = \{a_1, a_2, \dots, a_r\}$ is a finite part of (S) and $a_0 \in (S)$ then, from the associativity, $\oplus P$ means $a_1 \oplus a_2 \oplus \dots \oplus a_r$. So, $a_0 \oplus P = a_0 \oplus a_1 \oplus a_2 \oplus \dots \oplus a_r$. Such an operator defines a partial order relation \leq_\oplus over \mathbb{S} , by $x \leq_\oplus y$ if and only if $x \oplus y = x$. We assume that \mathbb{S} has a greatest element e_\oplus , such that $x \leq_\oplus e_\oplus$ for every $x \in \mathbb{S}$. Hence (\mathbb{S}, \oplus) is an Abelian idempotent semi-group with e_\oplus as identity element for \oplus .

Theorem 3.5 ([Tel88, Tel04]) *A wave can be used to compute an infimum.*

Self-Stabilizing Infimum Computation in a ϱ -ball. In order to add a initializing step, we assume $\delta = \varrho + 1$. We consider the following problem: at time $C_{U\delta}$ each register $p.v_0$ is initialized during the critical section $\ll CR2 \gg$, precisely when the register $p.r$ takes the value $UC\delta$. At the end of each phase $\Phi_U = [C_{U\delta}, C_{U\delta+\delta-1}]$, each process p needs to know the infimum of the registers $q.v_0$ of every q in $V(p, \varrho)$.

To reach the objective, we define for each process p two added registers $p.v_1$ and $p.v_2$. These two registers are initialized at the date $C_{U\delta}$ during the critical section $\ll CR2 \gg$, by the value $p.v_0$.

For $k \in \{1, 2, \dots, \varrho\}$, at the date $C_{U\delta+k}$, the action $\langle\langle CS1 \rangle\rangle$ is defined by:
 $p.v_1 := p.v_2$; $p.v_2 := p.v_0 \oplus \{q.v_{\omega(q)}, q \in \mathcal{N}_p\}$, with, if $q.r = p.r$ then $\omega(q) = 2$, and if $q.r = p.r + 1$ then $\omega(q) = 1$.

Proposition 3.6 *For $p \in V$ and $k \in \{1, \dots, \varrho\}$, at the date $C_{U\delta+k}$, both equalities hold:
(1) $p.v_1 = \bigoplus \{q.v_0, q \in V(p, k-1)\}$, and (2) $p.v_2 = \bigoplus \{q.v_0, q \in V(p, k)\}$.*

Proof.

At $C_{U\delta}$, any process p satisfies $p.v_1 = p.v_0$ and $p.v_2 = p.v_0$, it is the initializing step. Let \mathcal{A} the set of events in $\Phi_U = [C_{U\delta+1}, C_{U\delta+\varrho}]$, for which the proposition is not true.

Assume by contradiction that \mathcal{A} is not empty. Let (p, t) a minimal event in \mathcal{A} . Let $k \in \{1, 2, \dots, \varrho\}$ such that $(p, t) \in C_{U\delta+k}$. There exists t_0 such that $(p, t_0) \in \Phi_U$ and $(p, t_0) \rightsquigarrow (p, t)$. We have $p^{t_0}.v_1 = p^{t_0}.v_2 =$ and $p^{t_0}.v_2 = \bigoplus \{q.v_0, q \in V(p, k-1)\}$. This equality is true even if $k = 1$. Now, $p^t.v_2 = p.v_0 \oplus \{q^{t_q}.v_{\varphi(q)}, q \in \mathcal{N}_p\}$. From the minimality of the event (p, t) , the events (q, t_q) , where $t_q < t$, are not in \mathcal{A} and are in $[C_{U\varrho}, C_{U\varrho+\varrho-1}]$. So, $p.v_0 \oplus \{q^{t_q}.v_{\varphi(q)}, q \in \mathcal{N}_p\} = \bigoplus \{q.v_0, q \in V(p, k)\}$. We obtain a contradiction. \square

As a corollary, we obtain the expected theorem:

Theorem 3.7 *On the cut $C_{U\delta+\varrho}$, $p.v_2$ contains the infimum of the registers $q.v_0$ in the ϱ -ball centered in p , according to the phase $U = [C_{U\delta}, C_{U\delta+\varrho}]$.*

4 Applications

In this part, we show how to synchronize a self-stabilizing layer clock. The main clock defines a wavelet stream. Using the wavelet stream, we design with the slave clock a barrier synchronization at distance ρ . We then show how to use this layer clock to tackle efficiently many local synchronization problems at distance ρ .

4.1 Self-stabilizing Layer Clock

The idea is to manage the ϱ -wavelet stream. A clock organizes this stream. The wavelets are used to compute concurrently local infimum on each ϱ -ball. For each process p , once the infimum computed, a second clock defines a delay notion on the network. This delay is a total preordering useful to schedule the critical section enter of each process.

More formally, we define two clocks, the first clock C_1 (the master clock) and a second clock C_2 (the slave clock). The incrementing systems are respectively (χ_1, φ_1) and (χ_2, φ_2) . The behavior of the slave clock is scheduled by the first clock and a predicate *cond*. The predicate *cond* depends of the problem solved. To distinguish the two clocks, all the registers are subscripted by 1 or 2 respectively for the master clock and the slave clock. The predicates are superscripted by 1 or 2. For instance, the register of the master clock is denoted by r_1 and the register of the slave clock is denoted by r_2 , and the predicate $NormalStep_p^1$ is defined on the register r_1 of the process p . The predicate $NormalStep_p^2$ is defined on the register r_2 of the process p . We define in the same way WU_1 , WU_2 , and $WU = WU_1 \cap WU_2$. The stabilization of the layer clock is to guarantee $\Gamma \triangleright WU$.

When the system is stabilized, the schedule of the slave clock of a process p is defined in $\langle\langle CS2 \rangle\rangle$ by

$$NormalStep_p^2 \wedge cond \rightarrow \langle\langle CS2 \rangle\rangle; p.r_2 := \varphi_2(p.r_2)$$

Predicate *cond* is independent of the register $p.r_1$. It is this predicate which expresses the distance ϱ synchronization problem solved, while the procedures *Initialization* and *Computation* are scheduled by the wavelet and provide a preprocessing used by *cond*. The procedures *Initialization* and *Computation* depends on the problem to be solved to the layer clock. We give some instances for different problems later.

We define the predicate $NormalStep_p \equiv NormalStep_p^1 \wedge LocallyCorrect_p^2$.

Algorithm 2 (*SS – DC*) Self-stabilizing Layer Clock

Constant and variable:

\mathcal{N}_p : the set of neighbors of the process p ;

$p.r_1 \in \chi_1$; $p.r_2 \in \chi_2$;

Boolean Functions:

For clock $i \in \{1, 2\}$:

$ConvergenceStep_p^i \equiv p.r_i \in tail_{\varphi_i}^* \wedge (\forall q \in \mathcal{N}_p : (q.r_i \in tail_{\varphi_i}) \wedge (p.r_i \leq_{tail_{\varphi_i}} q.r_i));$

$LocallyCorrect_p^i \equiv p.r_i \in ring_{\varphi_i} \wedge \forall q \in \mathcal{N}_p, q.r \in ring_{\varphi_i} \wedge ((p.r_i = q.r_i) \vee (p.r_i = \varphi_i(q.r_i)) \vee (\varphi_i(p.r_i) = q.r_i));$

$NormalStep_p^i \equiv p.r_i \in ring_{\varphi_i} \wedge (\forall q \in \mathcal{N}_p : (p.r_i = q.r_i) \vee (q.r_i = \varphi(p.r_i)));$

$ResetInit_p^i \equiv \neg LocallyCorrect_p^i \wedge (p.r_i \notin tail_{\varphi_i});$

Common predicate:

$NormalStep_p \equiv NormalStep_p^1 \wedge LocallyCorrect_p^2;$

Actions:

$NA : NormalStep_p \longrightarrow$ **if** $p.r_1 \equiv \varrho - 1[\varrho]$ **then**
 Begin
 $NormalStep_p^2 \wedge cond \rightarrow$ $\ll CS\ 2 \gg ;$
 if $cond_1$ **then** $p.r_2 := \varphi_2(p.r_2)$
 Initialization ;
 End
 else *Computation*;
 $p.r_1 := \varphi_1(p.r_1);$

For clock $i \in \{1, 2\}$:

$CA_i : ConvergenceStep_p^i \longrightarrow p.r_i := \varphi_i(p.r_i);$

$RA_i : ResetInit_p^i \longrightarrow p.r_i := \alpha_i$ (**reset**);

Due to the lack of space, the proofs of Proposition 4.1 and 4.2 are left in the appendix.

Proposition 4.1 *The layer clock stabilizes to WU.*

Proposition 4.2 (No starvation) *Once stabilized, the clock C_1 increments infinitely often.*

4.2 Local Comparison in a ϱ -ball

In the network, when the layer clocks are stabilized, the delay between two processes according to the slave clocks defines a total preordering on the processes. Unfortunately this delay is a global notion. The problem is to find a condition such that for two processes in the same ϱ -ball, it is possible to calculate directly the delay with only the knowledge of the r_2 register values, and so to decide which process precedes the other according to the delay. To organize comparison between two processes lying in a same ball of radius equal to ϱ , it is sufficient to be able to compare the slave clock registers of any two processes at distance less than or equal to 2ϱ . By this way, in each ϱ -ball B , we will be able to define a total preordering among the processes in B by comparison of the values of the registers r_2 of the processes. Of course we want that this total preordering is the same than the preordering defined by the delay. In order to reach this objective, we extend the locally

comparability defined at one hop (refer to Section sub:unison) to the distance 2ρ . For the clarity, we must be more formal: A local order on a set χ is an antisymmetric and reflexive binary relation on χ . Let $\chi = \{0, \dots, K-1\}$ such that $K \geq 4\rho + 1$. Let a and b be two elements of χ . Let us assume that $d_K(a, b) \leq 2\rho$. Let us define now a local ordering \leq_l by : $a \leq_l b \Leftrightarrow_{def} 0 \leq \overline{b-a} \leq 2\rho$.

Lemma 4.3 *Let p and q be two processes satisfying $d(p, q) \leq 2\rho$. If $a = p.r_2$ and $b = q.r_2$ then the delay $\varrho_{p,q}$ is equal to $\overline{b-a}$ if $0 \leq \overline{b-a} \leq 2\rho$ and is equal to $-\overline{a-b}$ otherwise.*

Proof. Since $d(p, q) \leq 2\rho$, we have $\varrho_{p,q} \in \{-2\rho, \dots, 2\rho\}$. Moreover, $\varrho_{p,q} \equiv \overline{b-a}[K]$. Since, $K > 4\rho$, then $\varrho_{p,q}$ is equal to $\overline{b-a}$ if $0 \leq \overline{b-a} \leq 2\rho$, $-\overline{a-b}$ otherwise. \square

From this lemma, we access to the delay in each ρ -ball B . So our problem is solved. In the following section, we assume that $\rho = \rho + 1$, $K_1 = \rho K$ with $K_1 \geq C_G - 1$ and $K_2 \geq \max(4\rho + 1, C_G - 1)$. These assumptions ensure that the layer clocks are self-stabilizing, that the main clock is calibrate to defined a ρ -wave stream, and that delay defined by the slave clocks is computable at distance 2ρ with the only knowledge of the slave clock registers r_2 .

4.3 ρ -Local Resource Allocation

4.3.1 Problem Definitions

The *Resource Allocation* problem deals with resource sharing problems among the processes. The resource allocation allows processes to access resources, (i.e., their critical sections) concurrently, provided the resources are not conflicting with each other.

Definition 4.4 (Graph of Compatibility [CDP03]) *Let A be a set – sometime named the resource set–, let R be a reflexive binary relationship on A . We say that R is the compatibility relationship on A . If $(a, b) \in R$, then we say that a and b are compatible. If $(a, b) \notin R$ then we say that a and b are conflicting.*

The specification of the general resource allocation problem is defined as follows:

Safety: if a processor requests a resource in A to enter in critical section, then its request is eventually satisfied and it enters the critical section.

Fairness: In every execution, if two processes execute their critical section simultaneously, then both are using resources whose are compatible.

Most of the problem requiring coordination among process sharing some resources are particular instances of the graph of compatibility, and then, particular safety requirements. For instance, the following well-known problems are particular instances of the resource allocation problem:

Mutual exclusion: A is the set of processes and $R = \{(a, a), a \in A\}$. The safety condition is: *In every execution, no two processes execute their critical section simultaneously.*

Readers-Writer: A is the set of 2-uples $\{(p, r), p \in V, r \in \{read, write\}\}$, and R is defined by the safety condition: *In every execution, if two processes execute their critical section simultaneously, then both are executing a read operation.*

Group Mutual Exclusion: R is a equivalence relationship over a set of resources A . R is defined by the safety condition: *In every execution, if two processes execute their critical section simultaneously, then both are using resources in the same equivalence class.*

We now generalize the above requirements by limiting their effect to the conflict processes which are at ϱ -hops away of any given process p . Obviously, if $\varrho = D$, then the problem comes down to the above requirements. If $\varrho = 1$, then the set of processes which are conflicting with a process p is reduced to the neighboring processes of p . The most popular of this problems is the *dining philosopher* problem, also called the *Local Mutual Exclusion* (LME) problem.

ϱ -Local mutual exclusion: A is the set of processes and $R = \{(a, b) \in A^2, d(a, b) > \varrho \text{ or } a = b\}$

ϱ -Local Readers-Writer: A is the set of 2-uples $\{(p, r), p \in V, r \in \{read, write\}\}$, and R is defined by the safety condition: *In every execution, if two ϱ -neighboring processes execute their critical section simultaneously, then both are executing a read operation.*

ϱ -Local Group Mutual Exclusion: R is an equivalence relationship over a set of resources A . R is defined by the safety condition: *In every execution, if two ϱ -neighboring processes execute their critical section simultaneously, then both are using resources in the same equivalence class.*

The ϱ generic version called the **ϱ -Local Resource Allocation** (ϱ -LRA) is specified as follows: R is a general relationship over a set of resources A . The safety condition is: *In every execution, if two ϱ -neighboring processes execute their critical section simultaneously, then both are using resources whose are compatible.*

4.3.2 Self-Stabilizing Solutions

Due to the lack of space, the correctness proofs of this section are left in the appendix.

ϱ -LRA. Each process p maintains three registers $p.v : \Sigma$ where Σ is any data type, and the registers $p.res_1 : \chi_2$ and $p.res_2 : \chi_2 \times \Sigma$. The content of $p.v$ is the asked resource. Let us assume that a total order \preceq is defined on Σ . We reach the two fields of the register $p.res_i$ by $p.res_i.r$ and $p.res_i.v$ respectively, with $i \in \{1, 2\}$. We define a ϱ -local ordering on $\chi_2 \times \Sigma$ by:

$$(r, v) \blacktriangleleft (r', v') \Leftrightarrow (r <_l r') \text{ or } (r = r' \text{ and } v \preceq v')$$

Recall that $<_l$ is defined by the delay, which is a total preordering. This preordering being computable at distance ϱ . Define the associated ϱ -local infimum in the following way:

$$(r, v) \oplus (r', v') = \text{if } (r, v) \blacktriangleleft (r', v') \text{ then } (r, v) \text{ else } (r', v')$$

Define the macros *Initialization* and *Computation*, respectively by:

$$\begin{aligned} \text{Initialization} &\equiv p.res_1 := (p.r_2, p.v); p.res_2 := (p.r_2, p.v) \\ \text{Computation} &\equiv p.res_1 := p.res_2; p.res_2 := (p.r_2, p.v) \oplus \{q.res_{\omega(q)}, q \in \mathcal{N}_p\} \end{aligned}$$

with, if $q.r_2 = p.r_2$ then $\omega(q) = 2$ and if $q.r_2 = \overline{p.r_2 + 1}$ then $\omega(q) = 1$. The preprocessing of a local infimum designates a winner q in the ϱ -ball centered in p . It is important to see that this process is elected by p , and perhaps it is not elected by all the processes in the ϱ -ball centered in q . The condition *cond* depends of the solved problem, it is the disjunction: $(p.r_2, p.v) = p.res_2$ or *cond*₂. Where $(p.r_2, p.v) = p.res_2$ means that p is elected in the ϱ -ball centered in p , the condition *cond*₂ is there to raise concurrency, it depends of the solved problem. If p is not elected by the ϱ -local infimum calculation and *cond*₂ is true, the incrementation of the register may to be not wanted. The condition *cond*₁ of the incrementation is: $p.r_2 = p.res_2.r$.

ϱ -LME. Assume that each process p has an identifier denoted by $p.id$. The value of $p.id$ is in a total ordered set \mathbb{S} , for instance the integers. In fact, the identities must be only a 2ϱ -distance network coloring. That is to say that every node must be colored such that two vertices lying at distance less than or equal to 2ϱ have not the same color. We define for each process p the register $p.v$ as $p.id$. The couple $(p.r_2, p.id)$ is defined for each process p . The condition $cond_2$ is defined by *false* and thus $cond_1 \equiv \mathbf{true}$.

Definition 4.5 ϱ -LME has a fairness index of k , if in any computation, between any two consecutive critical section execution of a process, any other process can execute its critical section at most k times. The time-service of ϱ -LME is the maximal number of critical executions by other processes between two successive executions of the critical section by any process.

Our ϱ -LME algorithm is a barrier synchronization at distance ϱ . We deduce:

Proposition 4.6 For the ϱ -LME algorithm, the fairness index is equal to $\left\lceil \frac{D}{\varrho} \right\rceil$, and the service time is upper bounded by $\left\lceil \frac{n(n-1)}{\varrho} \right\rceil$.

From the definition of a phase [BPV04], and because K_1 is in $O(\varrho D) \subset O(D^2)$, K_2 is in $O(D)$, respectively. So, $K_1 K_2$ is in $O(D^3)$. Thus:

Proposition 4.7 (1) During one phase, the number of link-communications is equal to $2(\varrho + 1)|E|$, where $|E|$ is the number of edges in the network. (2) The stabilization time complexity of ϱ -LME algorithm is in $O(n)$ rounds, and (3) The space complexity of ϱ -LME algorithm is in $O(\log D)$.

Our solution for the ϱ -LME problem provide a good technique to reduce the service time and the fairness index of LME. The price to pay is the increase of the communications between processes.

ϱ -Group mutual exclusion. Let Σ be the set of resources. Let us assume that the preordering \preceq is defined on Σ –an arbitrary ordering, for instance a priority ordering. The binary relationship on A , defined by: $x \succ y$ iff $x \preceq y$ and $y \preceq x$, is an equivalence relationship. The equivalence classes are the groups. The set of groups is the quotient $\frac{\Sigma}{\sim}$. $p.v$ takes its values in $\frac{\Sigma}{\sim}$. An other way is to say that \preceq is an total ordering on the groups. To raise concurrency, if p asks a resource a and if the elected process at distance ϱ requests a resource in the same group, p enters in critical section. The predicate $cond$ is defined as: $\overline{p.v} = p.res.v$.

Note that we assume that there is no identity on the processes. However, we make the additional assumption that there is a total ordering on the resources. For instance, Local Mutual Exclusion problem is an instance of Group Mutual Exclusion where the resources are the processes. So, there is an ordering on the processes, which equivalent to define process identities.

ϱ -Readers-Writer. We assume that each process has an identity denoted by $p.id$. Each process has three possible requests: the process does not ask anything, the process asks to *read*, the process asks to *write*. This requests are symbolized respectively by N, R, W . In order to be able to compare two registers r at distance ϱ , we assume that $K \geq 4\varrho + 1$. For each ϱ -ball B , the local ordering \leq_l defines a total preordering on the registers r of processes in B . If a process p asks N or R then the register $p.v$ is initialized by the value F . If p asks W then $p.v$ is initialized by the value $Wp.id$. The ordering \preceq on Σ is defined by:

$$v \preceq v' \stackrel{\text{def}}{\equiv} (v = F \text{ and } v' = F) \text{ or } (v = WId \text{ and } v' = F) \\ \text{or } (v = W Id \text{ and } v' = W Id' \text{ with } Id \leq Id')$$

For each process p , the predicate $cond$ is defined by matching on $p.res$ as follows:

(r, F) when $r = p.r_2$	\rightarrow true
(r, Wid) when $r = p.r_2$ and p requests N	\rightarrow true
(r, Wid) when $r = p.r_2$ and $(p.r_2, p.id) = (r, id)$	\rightarrow true
--	\rightarrow false

5 Conclusion

We presented a self-stabilizing algorithm to solve the ϱ -wavelet scheme in arbitrary anonymous networks. Wavelets deals with coordination among processes which are at most ϱ hops away of each other. The proposed algorithm works under any (even unfair) daemon. Using the wavelet scheme, we described a self-stabilizing layer clocks protocol and showed that it provides an efficient device in the design of local coordination problems at distance ϱ , *i.e.*, ϱ -barrier synchronization and ϱ -local resource allocation (LRA) such as ϱ -local mutual exclusion (LME), ϱ -group mutual exclusion (GME), and ϱ -Reader/Writers. Some solutions to the ϱ -LRA problem (*e.g.*, ϱ -LME) allow to transform algorithms written assuming any ϱ -central daemon into algorithms working with any distributed daemon.

References

- [BDPV99] A Bui, AK Datta, F Petit, and V Villain. State-optimal snap-stabilizing PIF in tree networks. In *Proceedings of the Forth Workshop on Self-Stabilizing Systems (WSS'99)*, pages 78–85. IEEE Computer Society Press, 1999.
- [BPV04] C Boulinier, F Petit, and V Villain. When graph theory helps self-stabilization. In *PODC '04: Proceedings of the twenty-third annual ACM symposium on Principles of distributed computing*, pages 150–159, 2004.
- [BPV05] C Boulinier, F Petit, and V Villain. Synchronous vs. asynchronous unison. In *7th Symposium on Self-Stabilizing Systems (SSS'05), LNCS 3764*, pages 18–32, 2005.
- [BPV06] C Boulinier, F Petit, and V Villain. Toward a time-optimal odd phase clock unison in trees. In Springer-Verlag, editor, *Eighth International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS'06)*, Lecture Notes in Computer Science, 2006.
- [CDP03] S Cantarell, AK Datta, and F Petit. Self-stabilizing atomicity refinement allowing neighborhood concurrency. In Springer-Verlag, editor, *DSN SSS'03 Workshop: 6th Symposium on Self-Stabilizing Systems (SSS '03)*, volume 2704 of *Lecture Notes in Computer Science*, pages 102–112, 2003.
- [CHP71] P.J. Courtois, F. Heymans, and D.L. Parnas. Concurrent control with readers and writers. *Communications of the Association of the Computing Machinery*, 14(10):667–668, 1971.
- [Dij65] E.W. Dijkstra. Solution to a problem in concurrent programming control. *Communications of the Association of the Computing Machinery*, 8(9):569, 1965.
- [Dij68] E Dijkstra. *Cooperating Sequential Processes*. Academic Press, 1968.

- [Dij74] EW Dijkstra. Self stabilizing systems in spite of distributed control. *Communications of the Association of the Computing Machinery*, 17:643–644, 1974.
- [DIM97] S Dolev, A Israeli, and S Moran. Uniform dynamic self-stabilizing leader election. *IEEE Transactions on Parallel and Distributed Systems*, 8(4):424–440, 1997.
- [DNT06] P Danturi, M Nesterenko, and S Tixeuil. Self-stabilizing philosophers with generic conflicts. In *8th International Symposium on Stabilizing, Safety, and Security of Distributed Systems (SSS'06)*, 2006.
- [Dol00] S Dolev. *Self-Stabilization*. The MIT Press, 2000.
- [GGH⁺04] M Gairing, W Goddard, S T Hedetniemi, P Kristiansen, and A A McRae. Distance-two information in self-stabilizing algorithms. *Parallel Processing Letters*, 14(3-4):387–398, 2004.
- [GH99] M Gouda and F Haddix. The alternator. In *Proceedings of the Fourth Workshop on Self-Stabilizing Systems*, pages 48–53. IEEE Computer Society Press, 1999.
- [GHJT06] W. Goddard, S.T. Hedetniemi, D.P. Jacobs, and V. Trevisian. Distance- k information in self-stabilizing algorithms. In *The 13th International Colloquium On Structural Information and Communication Complexity Proceedings (SIROCCO'06)*, LNCS 4056, pages 349–356. Springer, 2006.
- [GMM04] E. Godard, Y. Métivier, and A. Muscholl. Characterizations of classes of graphs recognizable by local computations. *Theory of Computing Systems*, 37:249–293, 2004.
- [HC92] ST Huang and NS Chen. A self-stabilizing algorithm for constructing breadth-first trees. *Information Processing Letters*, 41:109–117, 1992.
- [Joh97] C Johnen. Memory-efficient self-stabilizing algorithm to construct BFS spanning trees. In *Third Workshop on Self-Stabilizing Systems*, pages 125–140. Carleton University Press, 1997.
- [Jou00] Yuh-Jzer Joung. Asynchronous group mutual exclusion. *Distrib. Comput.*, 13(4):189–206, 2000.
- [MN98] M. Mizuno and Nesterenko. A transformation of self-stabilizing serial model programs for asynchronous parallel computing environments. *Information Processing Letters*, 66(6):285–290, 1998.
- [NA02] M Nesterenko and A Arora. Stabilization-preserving atomicity refinement. *Journal of Parallel and Distributed Computing*, 62(5):766–791, 2002.
- [NS93] Moni Naor and Larry Stockmeyer. What can be computed locally? In *STOC '93: Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 184–193, New York, NY, USA, 1993. ACM Press.
- [Tel88] G Tel. Total algorithms. In Vogt FH (ed) Springer, editor, *Concurrency 88*, volume LNCS 335, pages 277–291. Springer-Verlag, 1988.
- [Tel04] G. Tel. *Introduction to Distributed Algorithms (Second Edition)*. Cambridge University Press, 2004.

A Self-stabilization of the layer clock

We apply the convergence stair method [Dol00].

Lemma A.1 *The predicates WU_1, WU_2 and WU are closed.*

Proposition A.2 *The clock C_1 stabilizes to WU_1 .*

Proof. Let $e = \gamma_1 \dots \gamma_k \dots$ be a maximal execution. Assume that e is finite. Then, the last state γ_l is a deadlock. So, the clock C_1 is stabilized, otherwise there should exist a process for which CA_1 or RA_1 is enable. We suppose now that e is not finite. The projection e_1 of e on the registers r_1 is an execution of the clock C_1 . If e_1 is finite, then in the last state C_1 is stabilized for the same reasons than above. If e_1 is not finite, then e_1 is an infinite execution of C_1 , so from [BPV04] there is a state which is in WU_1 . \square

Proposition A.3 *The clock C_2 stabilizes to WU_2 .*

Proof. Let $e = \gamma_1 \dots \gamma_k \dots$ be a maximal execution. We can assume from Proposition A.2 that $\gamma_1 \in WU_1$. So while C_2 is not stabilized, C_1 does not execute any action. So the projection e_2 of e on the registers r_2 is an execution of the clock C_2 . If e_2 is finite, then in the last state C_2 is stabilized otherwise there should exist a process for which CA_2 or RA_2 is enable. . If e_2 is not finite, then e_2 is an infinite execution of C_2 , so from [BPV04] there is a state which is in WU_2 . \square

From Proposition A.2 and Proposition A.3 we deduce the corollary:

Corollary A.4 *The layer clock stabilizes to WU .*

Proposition A.5 (No starvation) *Once stabilized, the clock C_1 increments infinitely often.*

Proof. Let $e = \gamma_1 \dots \gamma_k \dots$ be a maximal execution. We can suppose from Corollary A.4 that $\gamma_1 \in WU$. Assume that for a process p , action NA is executed only a finite number of time. Then the clock of each process is executed only a finite time, so e is finite. But in the last state of e , minimal processes according to the precedence relationship are enable, which is a contradiction.

While C_2 is not stabilized, C_1 does not execute any action. So the projection e_2 of e on the registers r_2 is an execution of the clock C_2 . If e_2 is finite, then in the last state C_2 is stabilized otherwise there should exist a process for which CA_2 or RA_2 is enable. If e_2 is not finite, then e_2 is an infinite execution of C_2 . From [BPV04] there is a state which is in WU_2 . \square

B Correctness Proofs of Subsection 4.3.2

ρ -LRA. In order to proof liveness and no lockout property, we lifts the main clock, using the lifting construction defined in section 3.2.2. We use the same notation, except that the register r becomes the register r_1 . As in Section 3.3, $\delta = \rho + 1$.

Lemma B.1 (No lockout) *In each phase $[C_{U\delta}, C_{U\delta+\rho}]$, there is at least one process which is elected and which increments.*

Proof. The set of processes is finite. So there is an infimum $(p.r_2, p.v)$ among the processes, according to the total order relationship \preceq . p is elected during the phase $[C_{U\delta}, C_{U\delta+\varrho}]$ and $p.r_2$ increments. \square

Lemma B.2 (Liveness) *Every process has the privilege infinitely often.*

Proof. It is sufficient to prove that each process p has the privilege at least one time. Assume that a process p has never the privilege. Let $Pot_p = \sum_{q \in V} \delta_{(p,q)}$ be the some of the delays from p to each q in V . This quantity is an integer, and from the assumption that p has never the privilege, Pot_p is strictly increasing. So Pot_p is not upper bounded. But this quantity is upper bounded by $|V|D$. This is a contradiction. We deduce the lemma. \square

ϱ -**LME.** From Section 4.3, both the no lockout and liveness properties are verified. It remains to show the safety property.

Lemma B.3 (Safety) *If the process p has the privilege, then no process at distance less than or equal to ϱ from p has the privilege simultaneously.*

Proof. Assume that p has the privilege in the phase $[C_{U\delta}, C_{U\delta+\varrho}]$, it enters critical section when its register satisfies $\tilde{p}.r_1 = U\delta + \varrho$. Any other process q at distance less than or equal to ϱ from p has not the privilege in the phase $[C_{U\delta}, C_{U\delta+\varrho}]$. So if $q \in B_\varrho(p)$ has the privilege simultaneously, it does not have the privilege in the same phase. The absolute value of the delay, according to the first clock, between the two processes is less than or equal to ϱ . But when p enters critical section and while p is in critical section: $\tilde{q}.r_1 \in \{U\delta + \varrho - \varrho, \dots, U\delta + \varrho + \varrho\} = \{U\delta, \dots, U\delta + 2\varrho\}$, and $U\delta + 2\varrho < (U + 1)\delta + \varrho$. We deduce that q can enter in critical section simultaneously if and only if $\tilde{q}.r_1 = U\delta + \varrho$, thus in the same phase as p , which leads to a contradiction. \square

ϱ -**Group mutual exclusion.** Again, from Section 4.3, both the no lockout and liveness properties are verified. The next lemma directly follows from the construction of *cond*:

Lemma B.4 (Safety) *If the process p and the process q have the privilege simultaneously and are at distance less than or equal to ϱ , the requested resources are in the same group.*