
Sara Sharifi Rad, Hassan Haghighi

Abstract—Due to important issues, such as deadlock, starvation, communication, non-deterministic behavior and synchronization, concurrent systems are very complex, sensitive, and error-prone. Thus ensuring reliability and accuracy of these systems is very essential. Therefore, there has been a big interest in the formal specification of concurrent programs in recent years. Nevertheless, some features of concurrent systems, such as dynamic process creation, scheduling and starvation have not been specified formally yet. Also, some other features have been specified partially and/or have been described using a combination of several different formalisms and methods whose integration needs too much effort. In other words, a comprehensive and integrated specification that could cover all aspects of concurrent systems has not been provided yet. Thus, this paper makes two major contributions: firstly, it provides a comprehensive formal framework to specify all well-known features of concurrent systems. Secondly, it provides an integrated specification of these features by using just a single formal notation, i.e., the Z language.

Keywords—Concurrent systems, Formal methods, Formal specification, Z language

I. INTRODUCTION

In the recent few years, concurrent processing has been almost everywhere in the computer world. In a concurrent system there exists a set of processes that execute concurrently. Also, each process interacts with other processes based on known approaches. Also, processes interaction is based on competition and/or cooperation. Threads which are in fact lightweight processes present a sample of cooperative processes existing inside a process. Cooperation of threads leads to the increase of concurrency, therupon multithreading concept is a basic context and extremely useful in concurrent systems [1], [3].

A concurrent system has many possible executions, and its behavior is usually not reproducible [2]. Consequently, the development of concurrent systems is a complex and error-prone task. Therefore, it is useful to specify, develop, and verify concurrent systems using formal methods. To develop a reliable concurrent system, it is significant to deduce relationship between properties of the concurrent system formally because the application of formal methods to the specification of systems is expected to increase the level of confidence in correctness of final programs [5]. In this way, formal methods have been long distinguished about the requirement to formally examine concurrent systems and provide an unambiguous description of these systems [4].

So far several formal specifications of concurrent systems have been presented by various methods and languages (e.g., VCD [14], TLZ [18] and Petri Net [21]). However, many aspects of concurrent systems, such as dynamic process creation, scheduling and starvation, have not been formally specified yet. Also, some other features have been specified partially and/or have been described using a combination of several different formalisms and methods whose integration needs too much effort. In other words, a comprehensive and integrated specification that could cover all aspects of concurrent systems has not been provided yet.

In this paper, we propose a comprehensive framework in order to formally specify all important features of concurrent systems, including Dynamic process creation, Multi-threading, Communication, Scheduling, Mutual exclusion, Deadlock, and Starvation using a single notation, i.e., the Z language, which provides us with mathematical techniques needed for specifying, verifying, and refining specifications into code formally. Thus, this paper makes two major contributions: firstly, it provides a comprehensive specification of concurrent systems covering all of their well-known features. Secondly, it provides an integrated specification of these features using a single formal notation, i.e., the Z language.

The paper is organized as follows: in section 2, we review related work. In section 3, a brief survey of formal methods and the Z language is presented. In section 4, we present our approach to specify concurrent systems. Finally, we conclude the paper in section 5.

II. RELATED WORK

In this section, we point to some related work. As can be seen in Table I, different methods and languages have been so far used to specify various features of concurrent systems. Also, these works do not cover all major aspects of concurrent systems.

Most of existing approaches of concurrent Z specifications have placed emphasis on the use of additional formalisms such as temporal logic, TLA and CSP [9–12]. Also, in some papers the behavioral and coordination aspects of concurrent systems are described by combining CCS and Temporal logic and/or GCCS [13], [14]. In this paper, we are going to specify all important aspects of concurrent systems fully based on the Z notation alone.
on concurrent systems have been partially proved; we now show this formalism could operate successfully to model all well-known features of concurrent systems.

IV. FORMAL SPECIFICATION OF CONCURRENT SYSTEMS PROPERTIES

As it has been shown in Table 1, some important aspects of concurrent systems, such as starvation, multi-threading and dynamic thread creation have not been specified yet. In addition, some cases have been specified in a way that is not related to the concurrent system exclusively; for example, the specification of scheduling has been presented for real-time systems not for concurrent systems.

In this section, we propose our comprehensive framework for formal specification of concurrent systems. The first step to achieve the above goal is the presentation of informal specification. More precisely, we provide useful definitions of concurrent system features in part A and then present the related formal specification in part B by referring to associated definitions in part A.

A. Principles of concurrent systems

Presented definitions in this section are derived from the features of concurrent systems [1], [2], [5], [15], [20], [21], [24], [26]–[30]:

Definition 1: Concurrency

A concurrent system is a collection of active entities that execute at the same time and interact with each other during their life cycle.

According to Definition 1, concurrent systems are composed of different components called active entities. The innuendo of active entity is process or thread. Each process has a unique name and independent address space. The process life cycle includes creation, scheduling and termination. If at a moment, more than one process is working, then we have indeed concurrency. Processes are execution units which can act in a concurrent manner if they interact with each other in a way that their executions overlap in time and/or there exists a combination of interleaving and overlapping.

A combination of these modes is shown in Fig. 1. In this figure, operations of process I and process II interleave in Time 1, and two operations I′ and I′ overlap in Time 2 because the second operation of process II (i.e., II′) is started before the second operation of process I (i.e., I′) is completed.

![Fig. 1 Combination of interleaving and overlapping](image-url)
Definition 2: Synchronization
Concurrent introduces the need for communication between executing processes; many resources may be shared between processes and threads in a concurrent system. Then the system requires a means to synchronize their operations.

According to Definition 2, in concurrent systems, both processes and threads need to synchronize among them in order to cooperate effectively when sharing resources or exchanging information. Related to this definition, there is the concept of critical section, a code segment in a process that accesses shared resources; these resources may be also accessed by other processes. Only one process must access its critical section at a time [26]. A solution to the critical section problem is mutual exclusion.

Definition 3: Coordinator
System resources are maintained and managed by a resource manager, called coordinator.

According to Definition 3, if a process in a concurrent system wants to access a resource, it must send a request message to the coordinator. A key word in concurrent systems is sharing: during execution, a resource, such as processor, memory and network, may be shared by various concurrent processes. It means that processes will compete for the resource. Thus, the shared resource must be protected by locking protocols. On the other hand, using the coordinator is one of the locking protocols that ensures mutual exclusion (refer to Definition 2) for concurrent executions.

Definition 4: Dynamic Thread Creation & Multi Threading
Each process, during its execution, can create several threads in own address space.

According to Definition 4, concepts multithreading and dynamic thread creation are taken. Threads of a process share their parent process address space. Unlike processes, threads do not have their own private address space, but share the state and global variables of a process together with other threads. These aspects have not been yet described formally in the literature.

Definition 5: Non-determinism
A program is non-deterministic if for at least one input, it produces more than one output and/or exhibits more than one behavior [24].

According to Definition 5, concurrent systems inherently exhibit non-deterministic behavior [26]. For example, when several processes compete for the same resource, non-deterministic effects appear [5]. In [24] the notion of multi-schema is defined as a tool for the specifier to specify non-determinism in Z explicitly. In this paper we use the same notation for modeling non-deterministic explicitly.

Definition 6: Communication
Processes need to communicate by passing data between them. Processes can be communicated in two ways: by shared variable or message passing.

When processes communicate by shared variables, one process “writes” into a variable that is “read” by another process, and when processes communicate by message passing, processes are assumed to share a communication network and exchange data in messages via “send” and “receive” primitives. Communication by message passing can be either synchronous or asynchronous.

In synchronous communication, communication happens only if the receiving process is waiting for the communication; this is termed a rendezvous. In asynchronous point to point message communication, a process sends a message to another process by placing the message in a location of network (Unlike the synchronous communication which uses networks as communication media, the asynchronous communication saves messages into networks); a location is an empty space in the network to hold the message. In an asynchronous communication, it is assumed that each network has an unlimited amount of location so that any number of messages may be placed in the network [15], [26].

Definition 7: Scheduling
During the execution of concurrent systems, fairness must be guaranteed by applying appropriate scheduling.

A scheduling policy is fair if it gives every process that is not delayed chance to proceed. On a single processor system, a scheduling policy is fair if it is unconditionally fair for processes that are not delayed, whereas, on a multi processor system, a scheduling policy is fair if it is unconditionally fair for parallel execution of processes. To specify scheduling in multi processor systems, we use Gang scheduling [20, 30] as a typical coscheduling approach that is widely used in concurrent systems. According to this scheduling strategy, a running process does not run forever; it eventually moves to the ready status, giving other processes the chance to proceed.

Definition 8: Standstill
A concurrent system is at the standstill state if no forward progress is being made.

Deadlock and livelock situations create standstill conditions for concurrent system [27], [29]. Deadlock is the most common problem in concurrent systems, and livelock term usually connotes Starvation and Infinite Execution. The relationship of these concepts is shown in Fig. 2.
Unlike livelock which may occur for one or more processes, deadlock always occurs for more than one process. Thus, in a deadlock status, whole of system is impaired while in a livelock status, only the current process(es) is (are) impaired. Further explanations are given in the next definitions.

**Definition 9: Deadlock**
Deadlock is a situation where two or more processes cannot proceed, because they are all waiting for another process to release some resources.

According to Definition 9, deadlock may occur when each process is waiting for another process to perform an operation. In many concurrent systems, the cost of deadlock avoidance is often considerable. Thus, these systems ignore the problems of deadlock until they enter a deadlock state. On the other hand, the occurrence of a deadlock can cripple part of a concurrent system [28]. Consequently, in this article we will consider two approaches to deadlock [27], [28] namely avoidance and detection & recovery approaches. The final decision will be taken in the implementation phase.

**Definition 10: Starvation**
A process with a non-zero cost may experience starvation.

Starvation may happen when one or more concurrent processes are blocked from gaining access to a resource. Consequently, such processes cannot progress [27], [29]. In this state the process status is Waiting continuously.

**Definition 11: Infinite Execution**
In a concurrent system, a process may execute forever. However, this process cannot progress.

Similar to the starvation status, in this state, the process cannot progress, but unlike the starvation status, in the infinite execution status, the process status exchanges between two statuses Running and Restart frequently.

**B. Formal specification of concurrent systems**
In this section, we propose a comprehensive and integrated framework for formal specification of all well-known features of concurrent systems reviewed in the previous subsection including dynamic thread creation, communication, scheduling, synchronization, deadlock, livelock, infinite execution and starvation using the Z language. It is worth mentioning that “Z/ves 2.1” has been used to validate the finally proposed specification. We now present our Z specification of concurrent systems step by step:

1. **Address Space, Message, PName**
   - The type of address spaces, messages and names of processes are specified by the above given types in Z.

2. **PT:**
   - Process, Thread

   Active entities in concurrent systems are processes and threads.

   - **Type Re:** Processor, Memory, Network
     - **Type Re** indicates type of resources in the system.

3. **Deadlock Approach:**
   - DetRec, AVO
   - Deadlock Avoidance: AVO
   - AVO
   - Deadlock Avoidance
   - AVO

   According to Definition 9, to obtain a comprehensive specification, we consider both Deadlock Detection & Recovery and Deadlock Avoidance approaches to deal with deadlock in this paper.

   - **Resource** is specified as follows:
     - **Resource**:
       - **type**: Type Re
       - **Location**: seq Message
         - **type** = Processor, Location = No

   - **Type Re** specifies the type of resources in the system.

   - **Location** in the network and memory as two main resources to hold the messages.

   - **Identifier** in Resource Schema specifies the type of the resource, and identifier **Location** is specified by a sequence of Message.

   - **Type OP** of process or thread operation is specified as follows:
     - **Type OP**:
       - Update, ReadOnly, Sender, Receiver

   According to Definitions 1 and 4, we use **Pr,Th** schema for Process and Thread specification as follows:

   - **Pr,Th**:
     - **Name**: PTName
     - **Pt**: PT
     - **NR**: Resource
     - **EM, IM**: Message
     - **Address**: Address Space
     - **ThreadsName**: PTName
     - **Type**: Type OP
     - **Status**: STATUS
     - **PreviousStatuses**: seq STATUS

   According to Definition 1, each process or thread has a unique Name. Thus, identifier **Name** indicates the unique name of the active entity. Identifier **pt** specifies the type of the active entity (Process or Thread) in the specification. This means that if **pt** is equivalent to Process, then all schema identifiers are related to process features; otherwise, all identifiers are associated to thread features.

   - **NR** specifies the set of resources requested by the process or thread right now. **EM** and **IM** show the set of Export and Import messages for each process or thread, respectively. If **pt** is equivalent to Process, then **threadsName** shows the set of names of threads which belong to the process. **PreviousStatuses** specifies the sequence of previous statuses of each processes or thread.
Now we write the initialization schema and all operation schemas of the concurrent system in turn:

\( \text{CS} \)

\( \text{processes' = } \begin{cases} \text{pr} \text{. Pr_Th} \end{cases} \)

\( \text{resources' = } \begin{cases} \text{pr} \text{. NR \& resources} \end{cases} \)

\( \text{coordinator' = Coordinator} \)

\( \text{communication' = Pr_Th \& Pr_Th} \)

\( \text{CT': Communication\_Type} \)

\( \text{DA': DetRec} \)

\( \text{DL\_s}\)

\( \text{P_Th = Process} \)

\( \text{p?: Pr_Th} \)

\( \text{new_t?: = Pr_Th} \)

\( \text{new_tn?: = Pr_Th} \)

\( \text{new_create!: Pr_Th} \)

\( \text{p?: status = Waiting} \)

\( \text{p?: = p} \)

\( \text{new_t?: = new_tn?: = #1 \_set} \)

\( p?: = \text{Process} \)

\( p?: = \text{Process} \)

\( p?: = \text{Process} \)

\( p?: = \text{Process} \)

\( \text{Create is an operation schema for creating a process in the system. In this schema, the input process (p?) will be created and added to the set of system processes.} \)

\( \text{CS'} \)

\( \text{processes' = } \begin{cases} \text{pr} \text{. Pr_Th} \end{cases} \)

\( \text{resources' = } \begin{cases} \text{pr} \text{. NR \& resources} \end{cases} \)

\( \text{coordinator' = Coordinator} \)

\( \text{communication' = Pr_Th \& Pr_Th} \)

\( \text{CT': Communication\_Type} \)

\( \text{DA': DetRec} \)

\( \text{DL\_s}\)

\( \text{P_Th = Process} \)

\( \text{p?: Pr_Th} \)

\( \text{new_t?: = Pr_Th} \)

\( \text{new_tn?: = Pr_Th} \)

\( \text{new_create!: Pr_Th} \)

\( \text{p?: status = Waiting} \)

\( \text{p?: = p} \)

\( \text{new_t?: = new_tn?: = #1 \_set} \)

\( p?: = \text{Process} \)

\( p?: = \text{Process} \)

\( p?: = \text{Process} \)

\( p?: = \text{Process} \)

\( \text{Create is an operation schema for creating a process in the system. In this schema, the input process (p?) will be created and added to the set of system processes.} \)
Release specifies abandonment of all the granted resources to a specific process or thread.

\[ \text{new}_{\text{nr}}! . \text{NR} = \text{new}_{\text{nr}}! . \text{NR} \]
\[ \text{p}! . \text{status} = \text{Restart} \]
\[ \text{r}! = \text{coordinator} . \text{Grant} \]
\[ \text{r} . \text{type} = \text{Processor} \]
\[ \text{new}_{\text{nr}}! = \text{Q} \cdot \text{p}: \text{processes} \cdot \text{p} = \text{p}? ; \text{p}. \text{NR} \& \text{r}! \]
\[ \text{new}_{\text{nr}}! = \text{Q} \cdot \text{p}: \text{processes} \cdot \text{p} = \text{p}? ; \text{p}. \text{NR} \& \text{r}! \]
\[ \text{new}_{\text{nr}}! = \text{Q} \cdot \text{p}: \text{processes} \cdot \text{p} = \text{p}? ; \text{p}. \text{NR} \& \text{r}! \]
\[ \text{processes}' \]
\[ = \text{p}: \text{processes} \cdot \text{p} \in \text{p}? \]
\[ \text{processes}' \]
\[ = \text{p}: \text{processes} \cdot \text{p} \in \text{p}? \]
\[ \text{processes}' \]
\[ = \text{p}: \text{processes} \cdot \text{p} \in \text{p}? \]
\[ \text{processes}' \]
\[ = \text{p}: \text{processes} \cdot \text{p} \in \text{p}? \]

When several processes compete for the same resource, non-deterministic effects appear since there may exist more than one process which can acquire a specific resource at the same time. Thus, the notion of multi-schema is used for specifying ND-Req.

Assign Resource includes the schema “ND-Req” above to complete the specification of resource allocation to a process existing in the resource queue.

SinScheduling and CoScheduling below are scheduling schemas for single-processor systems and multi-processor systems, respectively:

\[ \text{ND-Req} \]
\[ \text{new}_{\text{nr}}! . \text{NR} = \text{new}_{\text{nr}}! . \text{NR} \]
\[ \text{processes}' \]
\[ = \text{p}: \text{processes} \cdot \text{p} \in \text{p}? \]

According to Definition 7, fairness will be guaranteed by a suitable scheduler in the implementation phase, not in the specification stage.
According to Definition 8 and Fig. 2, SLI schema specifies Standstill-Livelock-Starvation state. According to Definition 10, if all previous statuses of a process are Waiting, then the process status is starvation.

According to Definition 8 and Fig. 2, SLI schema specifies Standstill-Livelock-Infinite execution. Now according to Definition 8, livelock situation is specified as follows:

CircularCondition checks deadlock possibility in a subset of processes. The output of this schema is either Yes or No.

CircularCondition

Synchronization

DeadLock_Recovery

Synchronization

p_loop?: p r Pr_Th
p?: & Pr_Th
p: Waiting
p: Waiting
p: Waiting
p: Waiting
p: Waiting
p: Waiting
p: Waiting
If deadlock approach is detection & recovery, then it is
resolved by killing a process or thread existing in the
detected cycle randomly; hence, we used the notion of multi-schema
when specifying DeadLock_Recovery.

According to Definition 6, in the asynchronous message
passing, a message can be placed on a location of the network,
provided there is some empty space in the network to hold the
message; it is assumed that each network has an unlimited
amount of space. Operation schemas As_Send_Me and
As_Receive_Me below specify sending and receiving
messages operations, respectively.

According to Definition 6, in shared memory systems,
write and read on shared variables; these operations are similar
In Synchronous SeAndRe schema specifies synchronous
message passing. According to Definition 6, in the
synchronous message passing, the sender process delays until
the receiving process is ready to receive the message.
Messages do not have to be saved in a location of the network.

According to Definition 6, in shared memory systems,
processes/threads communicate together using two operations
write and read on shared variables; these operations are similar
to send and receive in the asynchronous communication.
V. CONCLUSION AND FUTURE WORK

In this paper, at first the well-known aspects of concurrent systems including dynamic thread creation, communication, scheduling, synchronization, deadlock, livelock, infinite execution and starvation were reviewed informally. Then, a comprehensive and integrated framework for formal specification of these aspects was provided using the Z formal specification language. The final specification has been validated using a well-known Z type checker, i.e., Z/evs 2.1. In summary, we can assert that this paper benefits from the following advances in comparison to related work in the literature:

1. This work covers all well-known aspects of concurrent systems whereas some features of these systems, such as dynamic process creation, scheduling, and starvation, have not been specified formally yet.

2. Some other features of concurrent systems have been so far specified partially and/or have been described using a combination of several different formalisms and methods whose integration needs too much effort; see a detailed description in section 2; however, the formal specification proposed in this paper is fully based on a single formal specification language, i.e., the Z notation.

3. Unlike many other approaches in the literature, the work of this paper brings non-determinism into the specification explicitly. As it can be found in the previous section, we used the notion of multi-schema whenever we had to specify non-deterministic behavior. According to the discussion given in [5], such a specification leads to a program which preserves all allowable behaviors of the specified concurrent system.

One of the most important aims of specifying applications formally is to develop programs from formal specifications. We have chosen Z since it has an interpretation in Martin-Löf’s theory of types [31]. Therefore, as a future work, we are going to use this interpretation in order to translate our Z specification of a concurrent program into its counterpart in Martin-Löf’s theory of types and then drive a functional program from a correctness proof of the resulting type theoretical specification. In this way, we can provide a completely formal way to specify and develop concurrent systems.

REFERENCES


