Atomic Actions Based on Distributed/Concurrent Exception Resolution

A.Romanovsky
Computing Department, University of Newcastle upon Tyne, Newcastle upon Tyne, NE1 7RU, UK

The state of art in concurrent exception handling and resolution is discussed and a brief outline of all research in this area given. Our intention is to demonstrate that exception resolution is a very useful concept which should be used to facilitate joint forward error recovery in concurrent and distributed systems. To do this, several new arguments are considered. We regard resolution as reaching an agreement among cooperating participants of an atomic action. It is provided by the underlying system, which makes it unified and less error prone, and this is important for forward error recovery, complex by nature. We classify atomic action schemes into asynchronous and synchronous ones, and resolution implementations into centralised and decentralised ones. Another issue that we believe to be very important is about introducing atomic action schemes based on exception resolution into existing concurrent and distributed languages, which usually have only local (one-process) exceptions. We outline the basic approach and demonstrate its applicability by showing how resolution can be used in Ada83, Ada95 (for both concurrent and distributed systems) and Java. A discussion of ways to make this concept more object oriented and, with the help of reflection, more flexible and useful, concludes the paper.

Keywords: atomic actions, forward error recovery, exception handling, concurrent and distributed exception resolution, Ada83, Ada95, Java

1. Introduction

Fault-tolerant software [1] detects errors caused by faults and employs error recovery techniques to restore normal computation. Forward error recovery (mostly, exception handling schemes) is based on the use of redundant data and algorithms that repair the system by analysing the detected error and putting the system into a correct state. In contrast, backward error recovery returns the system to a previous error-free state without requiring any knowledge of the errors. The exception handling mechanism is a language/system feature allowing programmers to describe the replacement of the standard program execution by an exceptional execution when an occurrence of exception (i.e. anything inconsistent with the program specification) is detected (see [2] for a rigorous and thorough discussion). This mechanism is considered an essential part of any modern language (e.g. Ada95 [3], C++, Eiffel [4], Java [5]).

For any exception mechanism, *exception contexts* [2], i.e. regions in which the same exceptions are treated in the same way, have to be declared. Very often they are blocks or procedure bodies. Each context has a set of associated exception handlers; one of these is called when a corresponding exception is raised. There are different exception models. The termination exception model assumes that when an exception is raised, the corresponding handler completes the block execution. It is widely accepted that this model is more suitable for fault tolerant programming [1, 2]: it is adhered to in all mentioned languages. The resumption model assumes that the handler recovers the program state, and the program continues execution from the operation following the one that raised the exception. If there is no handler for the raised exception in the context or it is there but unable to recover the program, the exception is propagated. The exception propagation goes through a chain of procedure calls or of nested blocks. The appropriate handler is sought in the exception context containing that in which the handler was not found or was not able to recover the program.

2. Atomic actions. Concurrent exception resolution

Exceptions are much more difficult to handle and fault tolerance to provide in concurrent and distributed systems. The general approach to using exception handling in such systems proposed in [6] extends the well-known atomic action paradigm [1]. Atomic actions offer a general and sound basis for building fault tolerance schemes which allow processes to cooperate during recovery. [6] describes the main rules of cooperative recovery: involving all participating processes if any of them detects an error, and calling the features intended for recovery after the same error in all participants.

An atomic action is formed as a set of cooperative processes each of which participates in the action while executing its corresponding exception context. A set of exceptions is associated with each action. Each process participating in an action has a set of handlers for (all or part of) these exceptions. Participants enter the action by entering exception contexts. Their entries are asynchronous but they have to leave synchronously to guarantee that no information is smuggled to or from the action (this makes it easier to guarantee the main action properties). If an exception has been raised in a process, all action participants have to participate in the recovery, and the handlers for the same exception have to be called in all of them [6]. These handlers cooperate to recover the action. The participants can leave the action on three occasions. First of all, this happens if there have been no exceptions raised. Secondly, if an exception had been raised, and the called handlers have recovered the action. Thirdly, they can signal a *failure* exception to the containing action if an exception has been raised and it has been found that there are no appropriate handlers or that recovery is not possible.

A mechanism for *exception resolution* [6] is the essential part of concurrent exception handling since several independent exceptions can be raised at the same time, or several errors detected which are the symptoms of a different, more serious fault. [6] offered a solution which relies on using a *resolution procedure*: this resolves all concurrent exceptions and works out a generalised exception the handlers for which will be called in all action participants. The concept of the *exception tree* [6] is more appropriate than exception priorities for resolving these exceptions. This tree includes all exceptions associated with the action and imposes a partial order on them in such a way that a higher exception has a handler capable of handling any lower level exception.

None of the distributed and concurrent languages having exception handling features (e.g. Arche [7], Ada95, Ada83, Modula-3 [8], SR [9], Real-Time Euclid [10], Java) allows using atomic actions with forward error recovery. We will say that all these languages have a *local* exception handling.

3. Existing schemes

The paper [11] offers a general scheme for using an extended CSP to implement atomic actions with both forward and backward error recovery. To guarantee a synchronous exit together with the exception resolution, the authors [11] propose statically connecting all action participants in a virtual chain and synchronising them by rendezvous through this chain when they reach the end of the action (and, in particular, when exceptions are raised). This allows each process to receive the information about the exception from the 'left neighbour' process, to partially resolve the exception and to transmit the resolution result to the 'right neighbour' process. At the second step of this chain algorithm the last process in the chain finally resolves the exception and transmits the result to the 'left neighbour' process. This wave goes back to the left along the chain, and each process calls the appropriate handler for the same exception. We believe that though this scheme suggests some good ideas, it cannot be directly applied in practice, primarily because CSP is an experimental language. Moreover, the authors had to extend it by time-outs and exceptions. Another drawback seems to be that the scheme requires all participants to enter the action synchronously.

Some preliminary but very important steps were done for Ada83 in [12]. The authors discuss an Ada83 atomic action scheme which uses a service task (the action controller) having a set of nested statements **`accept`**, one for each participant. Each of them informs the controller of the code of the exception to be raised. Having received all codes, the controller raises the appropriate exception which propagates to all participants. We believe that this scheme is a promising approach that makes atomic actions practical. Still, the authors do not discuss the complete atomic action scheme or outline all peculiarities which should be discussed to allow the scheme to be used outright. Besides, we believe that this scheme is rather expensive (because of the repeated process synchronisation) and can be essentially improved (see [13] for a detailed analysis). Note that Ada95 is upwardly compatible for virtually all Ada83 applications and this scheme, in particular, can be directly used in Ada95.

Concurrent object-oriented language Arche [7] allows resolving failure exceptions signalled by several different implementations of the same class (view, in Arche terms), so this approach is not intended for coordinated recovery of concurrent cooperating processes or objects. The resolution procedure inputs all exceptions and returns one "concerted" exception to be handled in the context of the calling object. These procedures are application dependent because Arche uses parameterised exceptions and, as it is rightly pointed out in [7], the resolution tree is not applicable for these exceptions (one cannot know the parameters for the resolved exception).

The paper [14] discusses two very important issues: introducing concurrent exception handling into object oriented systems and a distributed resolution algorithm. Though this research is intended for the coordinated atomic (CA) action scheme [15] (a scheme that integrates conversations and transactions and allows both forward and backward recovery), it is quite general and can be used for any object oriented system, and so can the algorithm for resolving exceptions in any distributed system in which its assumptions are valid. This approach is essentially object oriented and exceptions are thought of on the class/object levels. The distributed resolution algorithm is less complex than the one briefly discussed in [6].

Basically, resolution implementations can be either decentralised or centralised depending on the way the synchronisation of participants with subsequent resolution is implemented. The two schemes in [11] and [12] are centralised and the location of the synchronising process is known statically: it is the head process for the first scheme and the controller for the second. The distributed scheme in [6] is essentially decentralised because all processes inform each other about their states and all of them resolve exceptions. Another distributed scheme in [14] is 'less' decentralised because the resolving process is found dynamically as the one having the biggest number of all processes in which an exception has been raised. The question involved is where to keep the resolution tree during the run-time. Depending on the implementation and on the resolution algorithm used, the tree can be kept either in all participants [6, 11, 14], or in one location (in the action manager [12]). Obviously, the resolution tree is kept where the resolution happens (or can happen): in schemes [6] and [11] all processes resolve exceptions (in [11] partially), but in schemes [14] and [12] only one site does so. The choice of the resolution implementation depends on application requirements and peculiarities, failure assumptions, underlying support, etc.

4. Why resolution is important. Resolution as agreement

We believe that the resolution mechanism is very important for concurrent and distributed systems and that it may be worthwhile to consider it as part of a future language or a distributed service (as a natural extension of the traditional exception mechanism). There are many reasons to support this and one of the intentions of this paper is to discuss them. The obvious reason is that several exceptions can be raised concurrently and this situation should be dealt with. Now we shall offer several arguments to explain why this may happen more often that one expects, why we cannot afford to ignore this situation and why the resolution should be included into the support rather than left to application programmers. We shall summarise the papers [6, 13, 14] and discuss some new topics.

First of all, since there are no perfect error detection tools, the latent period of an error is not negligible, and erroneous information can be easily spread within an atomic action; thus, several errors occurring concurrently in different processes can be the symptoms of a more serious fault. Secondly, very often there is a correlation between errors, so they happen over a very short period of time in different participating processes. With hardware-related errors, several processors or/and communication lines can be affected by the same bad conditions (in the case of a line this can affect all traffic going through it). With software design faults, as participating processes were designed cooperatively from the same specification, a cooperative misunderstanding can affect the design of all of them.

Another reason is that in practice it is impossible to interrupt all participants the moment one of them has raised an exception. And the probability that new exceptions are raised in other participants before they are informed about this exception is much higher in a distributed system. In distributed systems the overall hardware failure probability is very high and they are more difficult to program without design errors than centralised ones. Some languages (CSP, Ada83) simply have no features to interrupt or in any other way asynchronously inform the participating processes when one of them has raised an exception; as a result, they can be executed for quite a long period of time within which the errors caused by the same or by another fault could be detected in several of them.

Consider a situation when resolution is not used and concurrent exceptions are lost: handlers for only one of them are called in all participants. To cope with this situation, all handlers for all exceptions should be able to recover in situations either with or without these lost exceptions. But as the information about lost exceptions can be vitally important, all these handlers have to execute general error detection ('lost exception' detection), which can be very expensive. The original idea in [6] is that handlers should cooperate during recovery, but without resolution they will have to cooperate to detect errors before they can start recovery. After this, since each of them can find different errors, they are to cooperate to reach an agreement about the 'covering' error for which recovery actions should be started afterwards. It is clear that the approach in [6] tries to make recovery faster and more universal and programmers' jobs simpler because error detection which has been executed by all participants before exceptions have been raised is taken into account and because finding the covering exception is provided by the fault tolerance feature. It is obvious that, unlike backward error recovery, forward error recovery cannot be made transparent; it is inherently complex and application dependent, and traditionally only a basic exception scheme is used. Exception resolution makes this recovery much simpler. It should be regarded as reaching an agreement among action participants which is provided by the underlying support in a unified and automatic way. In Figure 1 we show two systems; in the first of these, application handlers have to repeat error detection and resolve the errors detected, but there is no need to repeat the error detection for the second one. Moreover, the underlying support executes resolution.

Now we would like to consider these arguments in detail. If exceptions can be lost, then, we believe, each participant has to find out (from within the handler) whether it had an exception which was lost and only afterward it can try to recover. Consider a situation when two participants raise `Fire_Alarm` and `Gas_Leakage` exceptions concurrently. It is obvious that we must not lose any of them and that the knowledge about both of them changes the recovery drastically. The same holds for situations when handlers for a less important exception are called in all participants and the information about a very important error is lost. This shows that all exceptions should be found when handlers are executed, and it is the responsibility of each handler.

Losing exceptions seems to be an idea that is not quite adequate for exception handling; the point is to pass as much information as possible between the normal and abnormal program states (e.g. parameterised exceptions have been introduced). Taking this idea to its extreme, we should use no exception names at all (but rather raise `The` exception), because handlers would have to detect the error(s) anyway.

Moreover, because all action participants have to be synchronised at the action end (this is the essence of all atomic action schemes) and it seems natural and not too costly to extend this final synchronisation by passing information about the detected errors and by an additional resolution stage.

5. Asynchronous and synchronous actions

There are two kinds of atomic action schemes: synchronous and asynchronous ones. In synchronous schemes, each participant has to either come to the action end or to find an error and to inform other participants of an exception; it is only afterwards that it is ready to accept information about their states. Asynchronous schemes do not wait for this but use some language feature to interrupt all participants when one of them has found an error. Generally speaking, resolution is required for both sorts of schemes. The Ada95 atomic action scheme [16] is essentially asynchronous, and it relies on ignoring all exceptions but one. So, there is no resolution: although several exceptions can be raised concurrently, only one of them will survive and the handlers for it will be called in all participants.

The appropriate approach should be chosen depending on the application, on the error which has been detected, on the failure assumption, etc. But the general scheme should, apparently, allow programmers to choose the more suitable approach.

Recovery and resolution in synchronous systems are much easier to provide than in asynchronous ones because each process is ready for recovery and is in a consistent state when handlers are called. Moreover, there is no need to program the abortion of nested actions for these systems because they have to be completed. Obviously, there is a risk that deadlocks can stop these systems, but we believe that cautious programming with an intensive error detection would make it possible not just to avoid this problem but make the subsequent recovery simpler. There is no wasting of time in asynchronous schemes but the corresponding features are not readily available in many languages and systems. Even when they are, they are usually very expensive: for example, many implementations of the asynchronous transfer of control in Ada95 will use the two thread model with the abortion and re-creation of one thread [17]. Moreover, they usually have complex semantics; it is more difficult to analyse, to understand and to prove programs which use these features. Very often some restrictions are imposed on the program segment that can be interrupted asynchronously, in an attempt to make the implementation less expensive. For example, Ada95 tasks cannot accept messages within this segment. One more difficulty with asynchronous schemes is that the abortion of nested actions is difficult to program. Some additional programming rules can make synchronous schemes better and decrease time waste (time-outs; assertions; checking invariants, pre- and post-conditions; etc.). This can allow an early detection of either the error or the abnormal behaviour of the process which has raised an exception and is waiting for the other processes.

The scheme in [11] is basically a synchronous scheme, although there is a proposal, based on a CSP extension, which allows an asynchronous scheme to be implemented (although this proposal is not complete, because if a process has no message input or output it cannot be informed by the broadcaster process). The paper [12] describes a synchronous scheme as Ada83 has no features to interrupt another process (other than just abort it).

6. Introducing exception resolution into existing languages

6.1. General approach

Now we would like to describe how exception resolution can be used in systems programmed in existing languages. This description should serve as a basis to be developed into a set of templates and conventions for application programmers to follow (we realise that this may be error prone, but there are some features which can help programmers to avoid mistakes: post- and pre-processors, libraries, syntax oriented editors, macro-processing). We do not want to introduce a new language construct for atomic action declaration: this would make the approach not feasible for existing languages. This is in line with some research reported recently (see, for example, schemes [12, 16, 18] which are intended for standard Ada83/Ada95). We believe now is the right time to map the fault tolerance approaches and schemes [1] that are very well researched but are not used in practice very often, onto practical, widely used existing languages. It seems to be one of the main flaws of the previous research in software fault tolerance that it is still rather theoretical and is applied to exotic systems and languages.

We would like to rely on the existing language features as much as possible and we will use their local exception mechanism as the base. We will rely on the general ideas [1, 6] about structuring concurrent software as a set of atomic actions. Let us describe how an atomic action scheme should be programmed. Each atomic action in this scheme is a dynamic entity consisting of a set of cooperative concurrent processes using their local exception handling. We follow the definition of the atomic action used in [1]: processes exchange information only among action participants. This restricts the behaviours of action participants: they must not interact with the outer processes and should leave the action synchronously. The scheme should guarantee either the synchronous exit of all processes from the actions when all of them have reached the end of exception contexts successfully, or the call of the handlers for the same resolved exception (even if several processes raise their exceptions). The scheme is to allow nested actions and exception propagation along nested exception contexts, corresponding to the chain of nested atomic actions. We adhere to the termination model, whereby handlers take over the duties of the processes participating in the action and complete it (either successfully or by signalling a failure exception to the containing action). The scheme should resolve concurrently raised exceptions. After such resolution, the handlers for the same exception are called in all participants, and they either cooperatively recover the action and fulfil the function specified by the action specifications or cooperatively signal a failure exception to the containing action.

We need a special synchronisation protocol in which all participants are to be involved and which may differ for different languages. All action participants start it by sending information about their states (successful completion, raising an exception). A special synchronising and resolving process (SR-process) collects all this information, resolves the exceptions which have been raised, and either raises the resolving exception in all participants or allows processes to leave the action. That means that the set of local exception contexts of all participants forms the atomic action. We assume that the SR-process has the resolution tree of all exceptions associated with given action and that all participants have local exception handlers for all exceptions from the tree. One general peculiarity is that we cannot allow any process to leave the action without informing the SR-process. The questions which have language-dependent answers are: passing the local exception to the resolving process, process synchronisation, raising the resolving exception. To demonstrate the applicability of the approach proposed we will outline several implementations.

6.2. Ada83

We consider the use of atomic actions in Ada83 as the first example of employing this approach. The scheme in [13] which we are going to describe is synchronous. One of the action participants is the head process, which executes nested statements **`accept`** when it finishes the action. Each of those is called by one of the other participants on finishing the action (with or without exception). An enumeration type containing the names of all action exceptions is to be declared for each action (e.g. `Action_A0_Exceptions_T`) and a value of this type is passed to the head process by each participant. When the head process has received information from all participants, it calls the resolution procedure to resolve exceptions and raises the covering exception. An interesting detail is that this exception is propagated by the

Ada83 run time to all action participants through the nested statements **accept**. For example, the head process for an action with three participants looks as follows:

```
 task P0 is
    entry RAISE_A0_P2(Exc_P2: in Action_A0_Exceptions_T);
    entry RAISE_A0_P1(Exc_P1: in Action_A0_Exceptions_T);
  end P0;
```

When it completes the action (in either way), it executes:

```
accept RAISE_A0_P2(Exc_P2: in Action_A0_Exceptions_T) do
    accept RAISE_A0_P1(Exc_P1: in Action_A0_Exceptions_T) do
        RESOLUTION_A0(Exc_P2, Exc_P1, Exc_C);
    end RAISE_A0_P1;
end RAISE_A0_P2;
```

The exception context for each participant (except for the head process) is as follows:

```
begin -- start of the exception context
   begin
     ... -- application code, participation in the action
      exception
        when Numeric_Error | Constraint_Error | Program_Error |
                     Storage_Error | Tasking_Error =>
             ... --  raising an action exception
   end; -- end of the additional block
exception
     when A => ... -- application recovery code
     when B => ... -- application recovery code
     when C => ... -- application recovery code
     when Universal_Exception => ... -- application recovery code
      ... -- raising the failure exception in the containing action
end; -- end of the exception context
```

An additional Ada83 block has been introduced to catch predefined exceptions. This scheme can be classified as a centralised one, with one resolving process keeping the resolution tree.

6.3. Ada95

Exceptions in Ada95 [3] are basically the same as in Ada83 and the approach above works for Ada95. But the new package `Ada.Exceptions` allows it to be simplified. Its function `Exception_Identity` returns the distinct identity (`Exception_Id`) of the exception raised. The modified scheme requires using two nested blocks to declare the exception context (identically to the basic scheme described) with the only handler `OTHERS` in the nested block. Exceptions are to be raised by Ada statement **raise**. In handler `OTHERS` `Exception_Id` of the raised exception is transferred to the head process as a parameter of the entry call. The resolution procedure manipulates the identities of exceptions raised, resolves them and raises the exception which will be handled by all action participants. The resolution procedure deals with the exception identities, which are kept in the resolution tree (this eliminates using exception values). The resolved exception is raised by procedure `Raise_Exception` from the same package. The handlers for all exceptions are to be declared in the second block of the exception context. An important detail is that an additional exception, `No_Exc`, should be declared and raised when a participant completes the exception block successfully. This modified scheme treats predefined exceptions and programmer's exceptions in the same way.

Further simplifications can be made using Ada95 protected types [3]. The parameterised protected type `SR_process` can be implemented as follows. It has two entries `Raise_Resolved` and `No_Exception`, which are to be called from the two corresponding handlers (`OTHERS` and `No_Exc`) of each action participant. The identities of the raised exceptions are collected in a list kept by `SR_process`. Procedure `Resolution` takes this list and, using the resolution tree, finds the resolved exception, which is assigned to variable `Resolved`. Note, that `Resolved` is equal to `Null_Id` provided all participants have raised exception `No_Exc`, in which case no exception is raised and the action completes successfully. An instance of this protected type is created for each action.

```
protected type SR_process(Participants_Number : Positive) is
        entry Raise_Resolved(E : in Exception_Id);
        entry No_Exception(E : in Exception_Id := Null_Id);
private
        procedure Resolution;
        entry Real_Raise(E : in Exception_Id := Null_Id);
        Finished : Integer :=0;
        Results : Results_T;
        Resolved : Exception_Id := Null_Id;
        Let_Go: Boolean := False;
end SR_process;

protected body SR_process is
        entry Raise_Resolved(E : in Exception_Id) when True is
        begin
                Finished:=Finished+1;
                Collect(E);
                if Finished=Participants_Number then
                        Resolution;
                        Let_Go:=True;
                end if;
                requeue Real_Raise;
        end Raise_Resolved;
        entry No_Exception(E : in Exception_Id := Null_Id) when True is
        begin
                Finished:=Finished+1;
                Results(Finished) := Null_Id;
                if Finished=Participants_Number then
                        Resolution;
                        Let_Go:=True;
                end if;
                requeue Real_Raise;
        end No_Exception;
        procedure Resolution is
        begin
                ... -- resolution procedure
```

```
        end Resolution;
        entry Real_Raise(E : in Exception_Id := Null_Id) when Let_Go is
        begin
                if Real_Raise'Count=0 then Let_Go:=False; Finished :=0;
                end if;
                Raise_Exception (Resolved);
        end Real_Raise;
end SR_process;
```

This is essentially a synchronous scheme; an analysis of Ada95 asynchronous atomic action schemes [16] shows that it is impossible not to lose some of the exceptions which are raised concurrently when Ada95's asynchronous transfer of control is used as the only means to implement an asynchronous scheme.

6.4. Distributed systems in Ada95

In this section we will give a sketch of programming atomic actions with exception resolution in distributed Ada95 programs (see Distributed Systems Annex [3]). The schemes above will not work in distributed systems because protected objects and task entries cannot be called remotely and `Exception_Id` cannot be passed between Ada95 partitions. We will modify the scheme proposed in Section 6.2 and make use of the fact that exceptions are propagated through remote procedure calls in Ada95. We assume that action participants (say, `P1`, `P2`, `P3`) reside on different partitions and we introduce a service partition, `SR_partition`, of category `Remote_Call_Interface` [3]. This partition (which is an Ada package) has a service procedure for each action participant:

```
package SR_partition is
        pragma Remote_Call_Interface;
        procedure P1_resolve (Exc_P1: in Action_A0_Exceptions_T);
        procedure P2_resolve (Exc_P2: in Action_A0_Exceptions_T);
        procedure P3_resolve (Exc_P3: in Action_A0_Exceptions_T);
end SR_partition;
```

Each package (including the service one) should be compiled with package `Pure` [3] containing types and data common for all action participants (action exceptions and the enumeration type):

```
package Action_A0 is
        pragma Pure;
        type Action_A0_Exceptions_T is (Exc_A, Exc_B, Exc_C, No_Exc, Failure);
        A, B, C : exception;
        Universal_Exception : exception;
end Action_A0;
```

Each task remotely calls its service procedure and passes the value of the exception this task is going to raise. There is a service task `SR_process` in this package that has a set of nested statements **`accept`** (similar to the scheme in Section 6.2). The resolution procedure is called from the most nested **accept** and raises the covering exception, which is propagated to all procedures: P1_resolve, P2_resolve, P3_resolve, and afterward to all action participants.

6.5. *Java*

The idea is again to rely on the local exception handling of Java [5]: the exception contexts (blocks try) of all action participants (threads) form the exception context of the atomic action. Each participant completes the execution of its exception context by throwing either an action exception it is going to raise, or a predefined exception noException (to inform about the successful completion). There is one service handler which catches all exceptions and calls service SRmethod, passing the exception as the parameter. SRmethod is a Java synchronised method and that guarantees that it is executed in an exclusive mode by only one thread (but if **wait** is executed, another thread can start execution). All but one threads are waiting on **wait** until the last one arrives, resolves the exception and notifies all of them (to let them proceed). Each of the threads throws the resolved exception:

```
        private exceptionActionA0 re;
private synchronized void SRmethod(exceptionActionA0 e)
                        throws excA, excB, excC, universalException {
        number--;
        if (!re.getClass().getName().equals ("noException")) collect(e);
        if number == 0 {
                        re = resolve();
                        number = before;
                        notifyAll();
        }
        else {
                try {
                        wait();
                } catch (InterruptedException ex) { ...
                }
        }
        if (!re.getClass().getName().equals ("noException")) throw re;
        }
}
```

Each block try in each action participant has to have blocks catch to catch all exceptions which are raised during the execution. These blocks should be included in the code, from the leaves of the resolution tree to the handler for the universal exception (the root), because the handler is searched for starting from the first block catch and the ancestor handler is called when a descendant exception is raised.

This scheme works for distributed Java systems which use remote method invocation [19].

7. Future research. Object orientation. Conclusions

Although several interesting ideas about ordering exceptions in other ways than trees (any partial order can be used, e.g. lattices) were mentioned in [6], we agree with the authors that the tree structure is the most suitable one. Since the paper was published, new evidence has been obtained to show that it matches object oriented system design. In particular, it works well for languages in which exceptions are classes: nearly all C++ and Java tutorials discuss the hierarchical structuring of exception classes. It remains to be seen how resolution procedures can be used for parameterised exceptions: although the paper [7] claims that the only way is for the former to be application-dependent, we believe that this problem can be solved by a proper choice of default values and by linguistic features which could allow dynamic binding based on procedure profile analysis and conformance rule checks. The problem of designing the corresponding handlers in the appropriate ways is getting more complex because these handlers are ordered, and any handler for any exception must be able to do all recovery which is done by the handlers for its descendant exceptions. On the other hand these handlers can be very complex because they should provide the cooperative action recovery. To make this approach practical and routine, clear and strict engineering rules should be described which explain how resolution should be used and how handlers should be designed. Although it is clear that software fault tolerance should be designed together with application software, the resolution tree can be built only when the designer is clear about the kinds of errors (and faults) that are to be tolerated, and the handlers can be implemented only after the tree has been built.

In the future using atomic actions with the exception resolution should be made more object-oriented. The right way is to regard actions as the instances of the special classes [14, 15]. The problems, which should be addressed are: the handlers and the resolution tree overriding and inheriting, the resolution tree extending, the tree and/or handlers reusing (e.g. what happens if we add a new exception, or override the old one; how can we re-order exceptions with minimum handler re-design; should we re-design and override all handlers on the tree path between the newly inserted one and the root; etc.).

We believe that the approach and implementations discussed can be used directly for introducing the CA action concept [15] into existing languages. In particular, for any centralised synchronous implementation the functions of the action manager can be extended for it to be the SR-process and to resolve concurrent exceptions.

Many object oriented systems allow reflecting upon their own behaviour and changing it by changing its representation. The main technique relies on using meta-object protocols [20], where each object is associated with a meta-object representing different aspects of the object behaviour. This opportunity to program specialised cooperating meta-objects can be of great use for exception resolution because if we can reflect upon raising exceptions and processes entering and leaving exception contexts, then we can move virtually the entire resolution, together with the synchronisation involved in the action completion, to a meta-level, which is a reasonable thing to do because the resolution protocol is not functional software. This can allow system programmers to design meta-object protocols relying on and choosing from resolution trees, lattices or other partial exception ordering; decentralised or centralised implementation of resolution; different agreement protocols (depending on the fault assumptions and the current system state). Meta-object protocols can change the resolution tree location or replicate it, for it to be used by several processes/sites, or to make the resolution procedure itself fault tolerant. Another protocol can deal with parameterised exceptions because, generally speaking, the entire object state is accessible from the meta-level and any required actual parameters can be picked from here. Besides, if there is an additional feature allowing asynchronous program interrupts, they can be used from within the meta-level, and synchronous schemes can be transformed into asynchronous ones (and vice versa). All this can be done transparently for functional application software and changed (adjusted) in the run-time.

Nowadays exception handling is obviously a very important part of a vast majority of practical languages. Many distributed and concurrent systems are programmed using these languages. A demand for concurrent/distributed exception handling, as a vital step in introducing forward error recovery into these systems, is recognised by many researchers because it makes this recovery simpler, uniform and less error prone. The main purpose of this paper is a thorough discussion of this subject. We believe that this feature should be regarded as part of the underlying support for future languages and systems.

Acknowledgements. This research is supported by DeVa Basic ESPRIT Project and my thanks go to all my colleagues in this project. I am grateful to A.Burns, A.Wellings and S.Mitchell (York University) for the fruitful discussions, criticism and sharing with me many ideas which helped me tremendously, and to A.Zorzo (Newcastle University) for helping me with Java scheme design.

References

[1] Lee, P. A. and Anderson, T.: 'Fault Tolerance: Principles and Practice' (Springer-Verlag, Wien - New York, 1990)

[2] Cristian, F.: 'Exception Handling and Tolerance of Software Faults' in *Software Fault Tolerance* Liu, M (Ed.) J. Wiley (1994)

[3] 'Information technology - Programming languages - Ada. Language and Standard libraries. ISO/IEC 8652:1995(E)'. Intermetrics, Inc., 1995

[4] Meyer, B.: 'Eiffel: The Language' (Englewood Cliffs, N.J., Prentice Hall, 1992)

[5] 'The Java Language Specification. Version 1.0 Beta.'. Sun Microsystems, Inc., 1995

[6] Campbell, R. H. and Randell, B.: 'Error recovery in asynchronous systems', *IEEE Trans. on Soft. Eng.*, 1986, SE-12, (8), pp. 811-826

[7] Issarny, V.: 'An Exception Handling Mechanism for Parallel Object-Oriented Programming: Towards Reusable, Robust Distributed Software.', *J. of Object-Oriented Programming*, 1993, 6, (6), pp. 29-40

[8] Nelson, G.: 'System Programming with Modula-3' (Prentice Hall, 1991)

[9] Huang, D. T. and Olsoon, R. A.: 'An exception handling mechanism for SR', *Computer Languages*, 1990, 15, (3), pp. 163-176

[10] Klingerman, E. and Stoyenko, A. D.: 'Real-Time Euclid: A Language for Reliable Real-Time Systems', *IEEE Trans. Softw. Engng.*, 1986, SE-12, (9), pp. 941-949

[11] Jalote, P. and Campbell, R. H.: 'Atomic Actions for Fault-Tolerance Using CSP.', *IEEE Trans. Softw. Engng.*, 1986, SE-12, (1), pp. 59-68

[12] Burns, A. and Wellings, A. J.: 'Real-time Systems and their Programming Languages' (Addison-Westley, New York, 1989)

[13] Romanovsky, A. B.: 'Practical exception handling and resolution in concurrent programs'. Computing Dept. University of Newcastle upon Tyne, 545, 1996

[14] Romanovsky, A., Xu, J. and Randell, B.: 'Exception handling and resolution in distributed object-oriented systems'. Proc. the 16th Int. Conference on Distributed Computing Systems, Hong Kong, 1996, pp. 545-553

[15] Xu, J., Randell, B., Romanovsky, A., Rubira, C., Stroud, R. and Wu, Z.: 'Fault tolerance in concurrent object-oriented software through coordinated error recovery'. Proc. the Twenty Fifth Int. Symp. on Fault-Tolerant Computing, USA, 1995, pp. 499-508

[16] Wellings, A. J. and Burns, A.: 'Implementing Atomic Actions in Ada95'. Department of Computer Science, University of York, YCS-263, 1996

[17] Burns, A. and Wellings, A.: 'Concurrency in Ada' (Cambridge University Press, 1995)

[18] Romanovsky, A. and Strigini, L.: 'Backward error recovery via conversations in Ada', *Software Engineering Journal*, 1995, 10, (8), pp. 219-232

[19] 'Java Remote Method Invocation Specification. Revision 0.9.'. Sun Microsystems, Inc., 1996

[20] Kiczales, G., Rivières, J. d. and Bobrow, D. G.: 'The art of the metaobject protocol' (The MIT Press, Cambridge, Massachusetts, 1991)

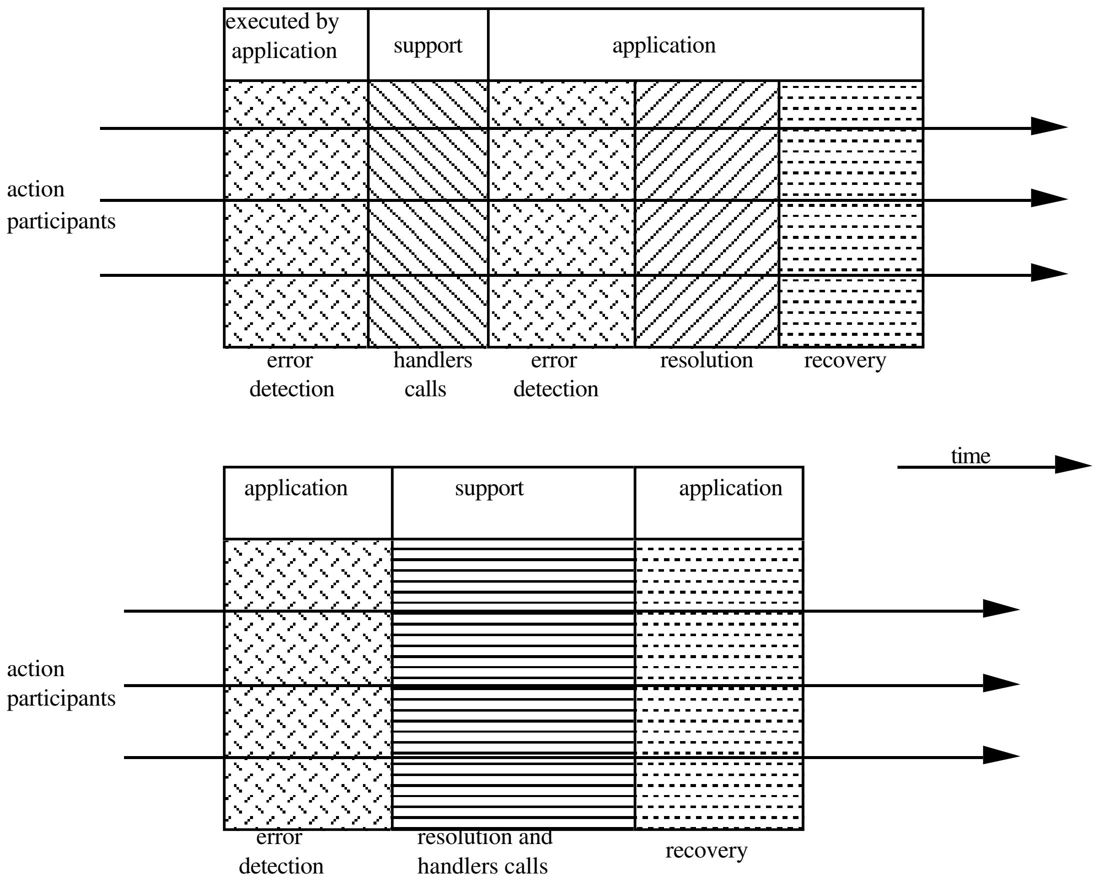

Figure 1. Action recovery for cases when resolution is executed by handlers (the first system) and by the underlying support (the second one)