Using the Strategy Design Pattern to Compose Reliable Distributed Protocols

Benoit Garbinato, Rachid Guerraoui
Laboratoire de Systemes d'Exploitation
Departement d'Informatique
Ecole Polytechnique Federale de Lausanne, Suisse

Using the Strategy Design Pattern to Compose Reliable Distributed Protocols*

Benoît Garbinato Rachid Guerraoui

Laboratoire de Systèmes d'Exploitation
Département d'Informatique
École Polytechnique Fédérale de Lausanne, Suisse
e-mail: bast@lse.epfl.ch

Abstract

Reliable distributed systems involve many complex protocols. In this context, *protocol composition* is a central concept, because it allows the reuse of robust protocol implementations. In this paper, we describe how the *Strategy* pattern has been recursively used to support protocol composition in the BAST framework. We also discuss design alternatives that have been applied in other existing frameworks.

1 Introduction

This paper presents how the Strategy pattern has been used to build BAST[1], an extensible object-oriented framework for programming reliable distributed systems. Protocol composition plays a central role in BAST and relies on the notion of protocol class. In this paper, we focus on the recursive use of the Strategy pattern to overcome the limitations of inheritance, when trying to flexibly compose protocols. In a companion paper [6], we have presented how generic agreement protocol classes can be customized to solve atomic commitment [10] and total order multicast [20], which are central problems in transactional systems and to group-oriented systems respectively. In [7], we also show how BAST allows distributed applications to be made fault-tolerant, by application programmers who are not necessarily skilled in reliability issues.

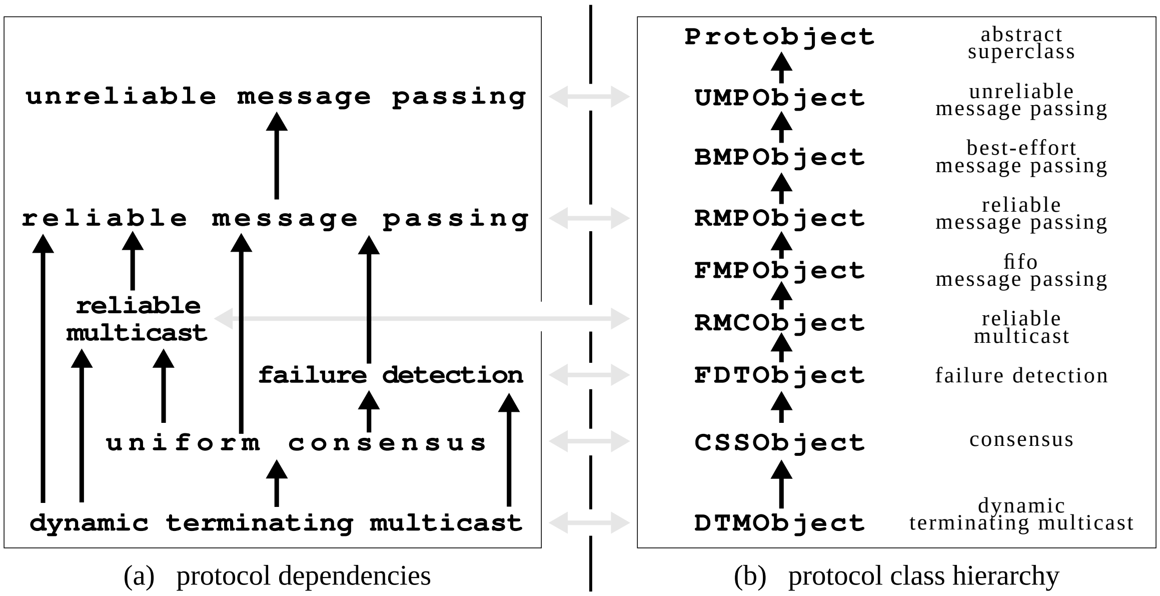
The BAST Framework

Building reliable distributed systems is a challenging task, as one has to deal with many complex issues, e.g., reliable communications, failure detections[2], distributed consensus, replication management, transactions management, etc. Each of these issues corresponds to some distributed protocol and there are many. In such a protocol "jungle", programmers have to choose the right protocol for the right need. Besides, when more than one protocol is necessary, the problem of their interactions arises, which further complicates programmers' task. The BAST framework aims at structuring reliable distributed systems by allowing complex distributed protocols to be composed in a flexible manner. For example, by adequately composing reliable multicast protocols with transactional protocols, BAST makes it possible to transparently support transactions on groups of replicated objects. It relies heavily on the *Strategy* pattern, which is recursively used to get around the limitations of inheritance as far as protocol composition goes. Our first prototype is written in Smalltalk [8] and is fully operational. It is currently being

*Partially supported by OFES under contract number 95.0830, as part of the ESPRIT BROADCAST-WG (number 22455)

[1] We named BAST after the cat-goddess of the Egyptian mythology: cats are known to survive several "crashes".

[2] A failure detector is a high-level abstraction that hides the timeouts commonly used in distributed systems [2].



(a) protocol dependencies (b) protocol class hierarchy

Figure 1: Protocols and Protocol Classes in BAST

used for teaching reliable distributed systems and for prototyping new fault-tolerant distributed protocols. Adding more and more protocol classes will help us to further test our approach. BAST has also been recently ported to the Java [9] programming environment. Performance is not yet good enough for practical application development, but we are currently working on performance evaluations and code optimization [7].

Overview of the Paper

Section 2 introduces the concept of protocol object as defined in BAST, and how it helps to structure distributed systems and to deal with failures. Section 3 discusses why inheritance alone is limited in supporting flexible protocol composition and presents how we applied the Strategy pattern to break these limitations. We also show how the Strategy pattern is transparently used in a recursive manner, and we present what steps have to be performed in order to extend BAST through protocol composition. Section 4 discusses various design alternatives, and compares our approach with other research works described in the literature. Finally, Section 5 summarizes the contribution of this paper, as well as some future developments in the BAST framework.

2 Protocol Objects

The BAST framework was designed to help programmers in building reliable distributed systems, and is based on protocols as basic structuring components. With BAST, a distributed system is composed of *protocol objects* that have the ability to remotely designate each other and to participate in various protocols. A *distributed protocol* π is a set of interactions between protocol objects that aim at solving *distributed problem* π. We use a `πObject` to name a protocol object capable of participating in protocol π, and we say that `πObject` is its *protocol class*. Each `πObject` provides a set of operations that implement interface protocol π, i.e., these operations act as entry points to the protocol. Abstract class `Protobject` is the root of the protocol class hierarchy.

With such broad definitions, any interaction between objects located on distinct network nodes is a distributed protocol, even a mere point-to-point communication. For example, class `RMPObject` implements a reliable point-to-point communication protocol and provides operations `rSend()` and `rDeliver()` that enable to reliable sending and receiving, respectively, of any object[3]; callback operation `rDeliver()` is redefinable and is said to be triggered by the protocol. Note that such a homogeneous view of what distributed protocols are does not contradict the fact that some protocols are more basic than others. Communication protocols, for example, are fundamental to almost any other distributed protocol.

Dealing with Failures. Because failures are part of the real world, there is the need for *reliable* distributed protocols, e.g., consensus, atomic commitment, total order multicast. Reliable distributed protocols are challenging to implement because they imply complex relationships with other underlying protocols. For example, both the atomic commitment and the total order multicast rely on consensus, while the latter is itself based on failure detections, on reliable point-to-point communications, and on reliable multicasts. In turn, reliable multicasts can be built on top of reliable point-to-point communications. Figure 1 (a) presents an overview of some distributed protocol dependencies.

In BAST, protocol classes are organized into a single inheritance hierarchy which follows protocol dependencies, as pictured in Figure 1 (b). Each protocol class implements only one protocol, but instances of some π`Object` class can execute any protocol inherited from π`Object`'s superclasses. Protocol objects are able to run several executions of identical and/or distinct protocols concurrently.

[3]We mean here any object that is *not* a protocol object. Allowing the sending of protocol objects across the network implies the solving of the distributed object migration problem. We did not address this issue in our framework yet.

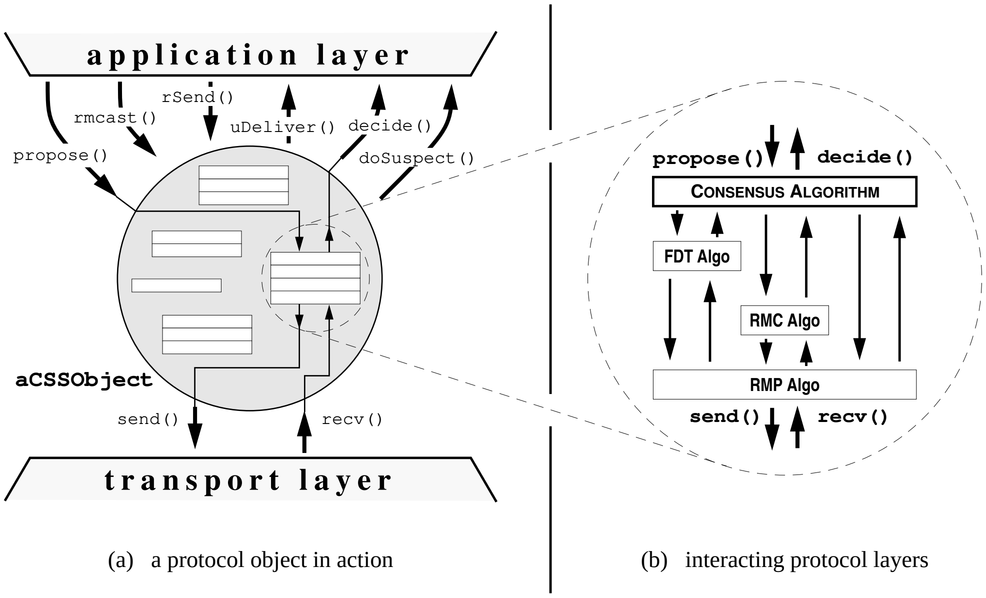
3 Strategy Pattern in BAST

Composing Protocols

With protocol objects, managing protocol dependencies is not only possible during the design and implementation phases (between protocol classes), but also at runtime (between protocol objects). This is partly due to the fact that protocol objects can execute more than one protocol at a time. In this context, trying to compose protocols comes down to answering the question "*How are protocol layers assembled and how do they cooperate?*".

Figure 2 (a) presents a runtime snapshot of `aCSSObject`, some protocol object of class `CSSObject` that implements an algorithm for solving the distributed consensus problem. The consensus problem is defined on some set σ of distributed objects as follows: all correct objects in σ propose an initial value and must reach agreement on one of the proposed values (the decision) [3]. Class `CSSObject` defines operations `propose()` and `decide()`, which mark the beginning and the termination of the protocol respectively [2]. Besides consensus, protocol object `aCSSObject` is also capable of executing any protocol inherited by its class, e.g., reliable point-to-point communications and reliable multicasts, as well as failure detections. In Figure 2 (a), `aCSSObject` is concurrently managing five different protocol stacks for the application layer, and issuing low-level calls to the transport layer. Focusing on the consensus stack, protocol composition means here to assemble various layers, each being necessary to execute the consensus protocol, into the protocol stack pictured in Figure 2 (b). The assembling occurs at runtime and creates a new stack each time the application invokes operation `propose()`.

Inadequacy of Inheritance Alone. With BAST, distributed applications are structured according to their needs in protocols: they are made of protocol objects, which act as distributed entities capable of executing various protocols. With this approach, it all comes down to choosing the right class for the right problem. We believe that inheritance is an



(a) a protocol object in action

(b) interacting protocol layers

Figure 2: Protocol Layers and Protocol Objects

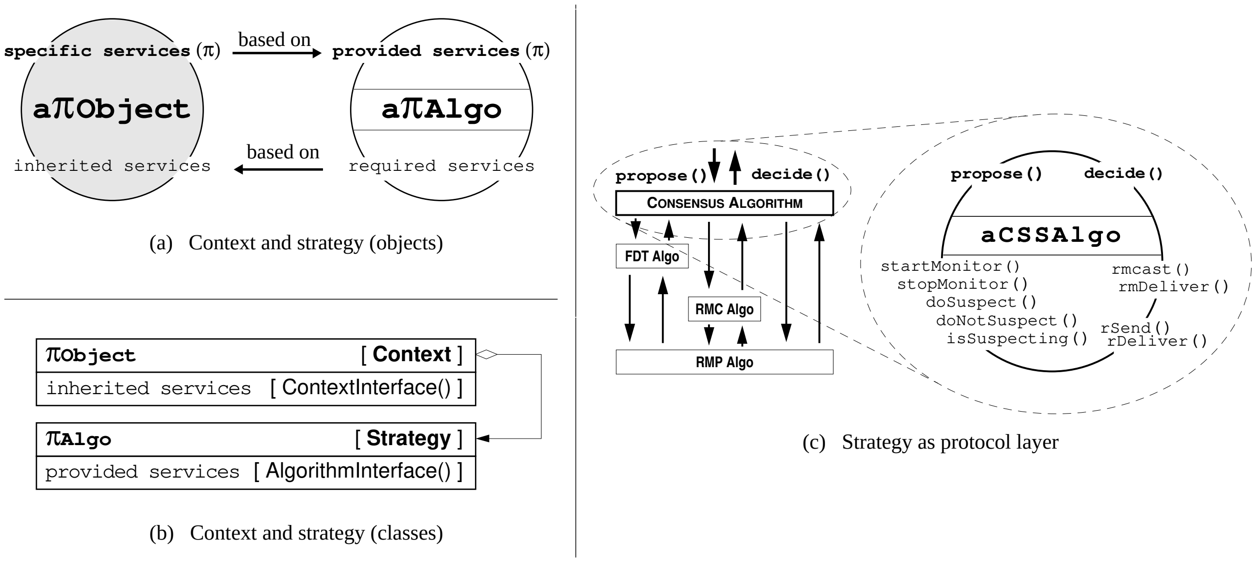
appropriate tool to achieve this: by passing appropriate arguments to protocol operations and by implementing callback operations, programmers have the ability to tailor generic protocol classes to their needs. However, we claim that inheritance alone is not sufficient as far as protocol composition goes, because it does not offer enough flexibility. For example, inheritance does not allow for the easy implementation of a new algorithm for some existing protocol, and then to use it in various protocol classes that are scattered in the class hierarchy. Furthermore, inheritance is not appropriate when it comes to choosing among several protocol algorithms *at runtime*. These limitations lead us to seek an alternative solution for flexible protocol composition.

Protocol Algorithms as Strategies

According to Gamma et al., the intent of the *Strategy* pattern is to "*define a family of algorithms, encapsulate each one, and make them interchangeable*" [5, page 315]. This is usually achieved by objectifying the algorithm [4], i.e., by encapsulating it into a so-called *strategy* object; the latter is then used by a so-called *context* object. Making each π`Object` protocol class independent of the algorithm supporting protocol π is precisely what we need to be able to compose reliable distributed protocols in a flexible manner.

In the BAST framework, strategy objects represent protocol algorithms and they are instances of subclasses of class `ProtoAlgo`. A `ProtoAlgo` subclass that implements an algorithm for solving problem π is referred to as class π`Algo`. In the Strategy pattern terminology, a protocol algorithm, instance of some π`Algo` class, is a *strategy*, and a protocol object, instance of some π`Object` class, is a *context*. A strategy and its context are strongly coupled and the application layer only deals with instances of π`Object` classes, i.e., it knows nothing about strategies.

Strategy/Context Interactions. Figure 3 (a) sketches the way protocol objects and algorithm objects interact. On the left side, protocol object aπ`Object` offers the services it inherits from its superclasses, as well as the new services that are specific to protocol π. The actual algorithm implementing protocol π is not part of aπ`Object`'s code; instead, the latter uses services provided by strategy aπ`Algo` (on the right side of Figure 3 (a)).



(a) Context and strategy (objects)

(b) Context and strategy (classes)

(c) Strategy as protocol layer

Figure 3: Strategy Pattern in BAST

Whenever an operation related to protocol π is invoked on aπ`Object`, the execution of the protocol is delegated to strategy aπ`Algo`. In turn, the services required by the strategy to run protocol π are based on the inherited services of context aπ`Object`. Such required services merely identify entry point operations to underlying protocols needed to solve problem π.

Each instance of class π`Algo` represents one execution of protocol π implemented by that class, and holds a reference to the context object for which it is running; any call to the services required by the strategy will be issued to its context object. There might be more than one instance of the same `ProtoAlgo`'s subclass used simultaneously by aπ`Object`. At runtime, the latter maintains a table of all strategies that are currently in execution for it. Each message is tagged to enable aπ`Object` to identify in which execution of what protocol that message is involved, and to dispatch it to the right strategy. Figure 3 (b) presents the relationship between classes π`Object` and π`Algo`, using a class diagram based on the Object Modeling Technique notation [19]. The correspondence between π`Algo` strategy objects and layered protocol stacks is pictured in Figure 3 (c): at runtime, each strategy object represents a layer in one of the protocol stacks currently in execution.

Consequences. The context/strategy separation enables the limitations of inheritance to be overcome, as far as protocol composition goes. One could for example optimize the reliable multicast algorithm and use it in some protocol classes, while leaving it unchanged in others. Protocol algorithms could even be dynamically edited and/or chosen, according to criteria computed at runtime; this feature is analogous to the dynamic interpositioning of objects. There is a minor compatibility constraint among different protocol algorithms in order to make them interchangeable: new algorithm class $\pi\mathtt{Algo}_n$ can replace default π`Algo` in protocol class π`Object` *if and only if* $\pi\mathtt{Algo}_n$ requires a subset of the services featured by π`Object`.

This approach also helps protocol programmers to clearly specify, for each protocol π, its dependencies with other protocols. One drawback of the Strategy pattern is the overhead due to local interactions between strategies and contexts. In distributed systems however, this overhead is small compared to communication delays, especially when failures and/or complex protocols are involved. More specifically, the time for a local Smalltalk invocation is normally under 100 μs, whereas a reliable multicast communication usually takes more than 100 ms when three or more protocol objects

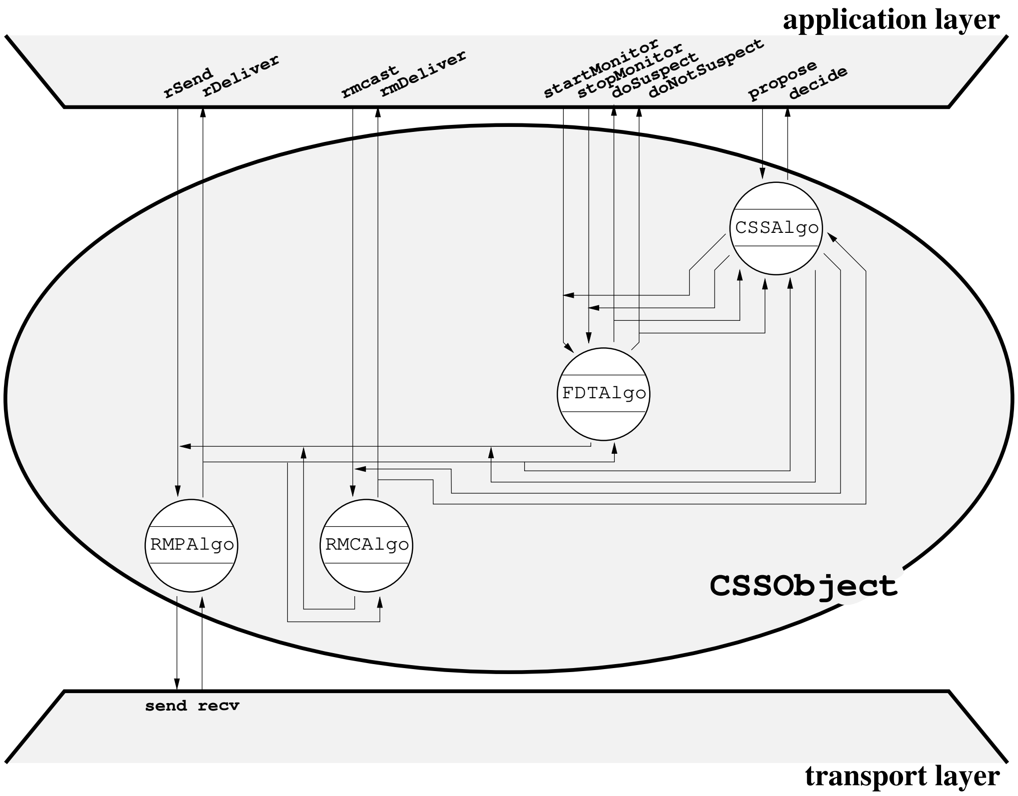


Figure 4: Recursive Use of the Strategy Pattern

are involved[4] (without even considering failures). The gain in flexibility clearly overtakes the local overhead caused by the use of the Strategy pattern.

Reliable Multicast: an Example

We now present how we implemented reliable *multicast* communications using the Strategy pattern. In BAST, class `RMCObject` provides primitives `rmcast()` and `rmDeliver()` that enable the sending and receiving, respectively, of a message `m` to a set of protocol objects referenced in `destSet`, in a way that enforces reliable multicast properties. The current implementation of class `RMCObject` relies on strategy class `RMCAlgo`.

Overview of the Protocol. The protocol starts when operation `rmcast()` is invoked on some initiator object $\texttt{aRMCObject}_i$, passing it a message `m` and a destination set `destSet`. In this operation, context $\texttt{aRMCObject}_i$ first creates a strategy $\texttt{aRMCAlgo}_i$, and then invokes operation `rmcast()` on it, with the arguments it just received. Strategy $\texttt{aRMCAlgo}_i$ builds message $\tilde{\texttt{m}}$, containing both `m` and `destSet`. It then issues a reliable point-to-point communication with each protocol object referenced in `destSet`; in order to do this, strategy $\texttt{aRMCAlgo}_i$ relies on inherited service `rSend()` of context $\texttt{aRMCObject}_i$. When message $\tilde{\texttt{m}}$ reaches $\texttt{aRMCObject}_t$, one of the target objects, operation `rDeliver()` is triggered by the protocol. Operation `rDeliver()` detects that $\tilde{\texttt{m}}$ is a multicast message and forwards it to $\texttt{aRMCAlgo}_t$, the strategy in charge of that particular execution of the reliable multicast protocol. When $\texttt{aRMCAlgo}_t$ receives $\tilde{\texttt{m}}$ for the first time, it re-issues a reliable point-to-point communication with each protocol object referenced in `destSet` (extracted from $\tilde{\texttt{m}}$), and then invokes `rmDeliver()` on its context $\texttt{aRMCObject}_t$, passing it message `m` (also extracted from $\tilde{\texttt{m}}$). This

[4] On a 10 Mbits Ethernet connecting Sun SPARCstations 20.

retransmission scheme is necessary because of the *agreement* property of the reliable multicast primitive, which requires that either all correct objects in `destSet` or none receive message `m` [2].

Recursive Use of the Strategy Pattern

When solving distributed problem π, one can strictly focus on the interaction between class π`Object` and class π`Algo`, while forgetting about how other protocols are implemented. In particular, all protocols needed to support protocol π are transparently used through inherited services of class π`Object`. Those services might also be implemented applying the Strategy pattern, but this is transparently managed by inherited operations of π`Object`. In that sense, BAST uses the Strategy pattern in a powerful *recursive* manner.

The recursive use of the Strategy pattern is illustrated in Figure 4. The latter schematically presents a possible implementation of protocol class `CSSObject` presented in Section 3, which enables the solution of the distributed consensus problem by providing operations `propose()` and `decide()`. In Figure 4, the gray oval is context class `CSSObject`, while inner white circles are various π`Algo` strategy classes (π being different protocols). Arrows show the connections between provided services (top) and required services (bottom) of each strategy class. Operations provided by class `CSSObject` are grouped on the application layer side (top). Each strategy class pictured in Figure 4 is managed by the corresponding context class in the protocol class hierarchy presented in Figure 1 (b).

Extending the BAST Framework

Basing the BAST framework on the Strategy pattern has the advantage of making it easily extensible. To illustrate this, we now present how we built `DTMObject`, a protocol class supporting the *Dynamic Terminating Multicast* (DTM) protocol [11] from existing contexts and strategies. The DTM protocol can be understood as a common denominator of many reliable distributed algorithms [12].

Overview of the Protocol. The protocol starts by the invocation of operation `dtmcast()` on an initiator object, passing it a message `m` and a set of protocol object references `destSet`. This invocation results in a reliable multicast of `m` to the set of participants objects. When message `m` reaches some participant, the protocol triggers operation `dtmReceive()`, passing `m` as argument. The participant object then computes a reply and returns it. Eventually, operation `dtmInterpret()` is triggered by the protocol on each non-faulty participant object, taking `replySet`, a subset of the participants' replies, as argument. The protocol insures that all correct participant objects get the same subset of replies, i.e., a consensus has been reached on that set.

Methodology for Extending BAST. A five steps methodology guides programmers in extending the BAST framework using the Strategy pattern. We illustrate each of these steps below, by presenting how the methodology was applied to the design of class `DTMObject`. Figure 5 summarizes the methodology.

1. Establish what services the new protocol class `DTMObject` provides, i.e., what operations are given to programmers wanting to use `DTMObject`; those operations are `dtmcast()`, `dtmReceive()` and `dtmInterpret()`.

2. Choose an algorithm implementing DTM and determine what services it requires, by decomposing it in a way that allows to reuse as many existing protocols as possible; those services are: consensus, failure detections, as well as reliable point-to-point and reliable multicast communications (see [11] for algorithmic details).

3. Implement the chosen algorithm in some `DTMAlgo` class; all calls to the above required services are issued to an instance variable representing the context object, i.e., an instance of class `DTMObject`.

1. Establish the DTM$_S$ specification of protocol DTM. (i.e., *provided* services).
2. Determine the services *required* by the chosen algorithm for DTM$_S$.
3. Implement that algorithm in a subclass of class `ProtoAlgo`. (Let's name it `DTMAlgo`).
4. Choose a subclass of class `Protobject` offering *at least* all services defined by step 2. (Let's take class `CSSObject`).
5. Subclass `CSSObject` into class `DTMObject`. The implementation merely "connects" *provided* and *required* services of `DTMAlgo`, to *specific* and *inherited* services of `DTMObject` respectively.

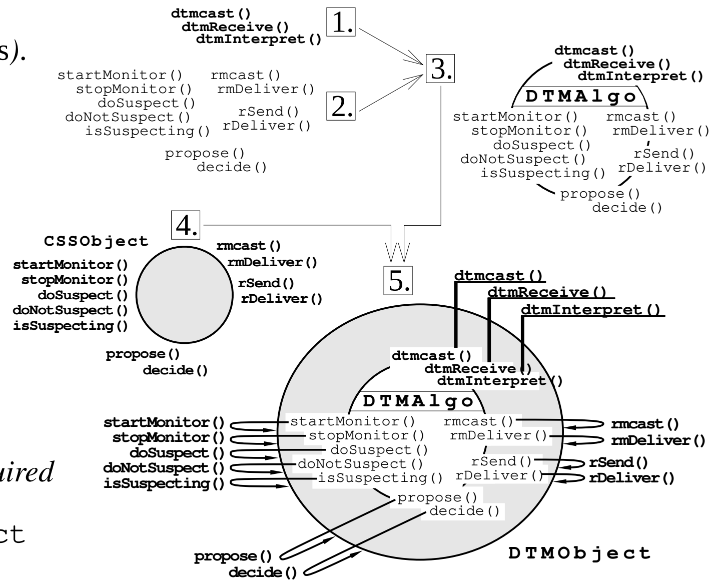


Figure 5: Extending BAST with Protocol Class `DTMObject`

4. Choose the protocol class that will be derived to obtain new class `DTMObject`; the choice of class `CSSObject` is directly inferred from step 2, since the chosen superclass has to provide *at least* all the services required by protocol DTM.

5. Implement class `DTMObject` by connecting services provided by class `DTMAlgo` to new DTM-specific services of class `DTMObject`, and by connecting services required by class `DTMAlgo` to corresponding inherited services of class `DTMObject`.

4 Design Alternatives

Our first implementation of BAST was not based on the Strategy pattern, i.e., distributed algorithms were not objects, and protocol objects were not capable of participating in more than one protocol execution concurrently. Furthermore, protocol composition was only possible through *single* inheritance[5].

[5]Remember that we used Smalltalk as implementation language for prototyping.

Because protocol objects are the basic addressable distributed entities in our approach, it is not possible to guarantee that there will never be more than one protocol execution involving each protocol object at a given time. For example, we cannot make sure that there will not be two concurrent multicast communications and/or transactions involving the same protocol objects. Allowing concurrency at this level is an essential feature. Moreover, as far as protocol composition is concerned, single inheritance is inadequate for offering a satisfactory degree of flexibility.

For all these reasons, we made BAST evolve through a second implementation of which the main goal was to overcome the limitations mentioned above. We now discuss some design alternatives that were considered in the process of implementing this second prototype of BAST, together with design issues that we studied from other existing frameworks described in the literature.

Multiple Inheritance and Mixins

Although our prototyping language does not offer multiple inheritance, assembling the various protocol layers through this code reuse mechanism is very appealing[6]. The idea is to make each protocol class π`Object` implement only protocol π, while accessing all required underlying protocols through unimplemented operations; each protocol class is then an *abstract* class and we usually say it is a *mixin class* or simply a *mixin*. Before being able to actually instantiate a protocol object, one first has to build a new class deriving from all the necessary mixins.

There are three major drawbacks with this approach. First, protocol classes are not more ready-to-use components: a fairly complex multiple subclassing phase is now required. As consequence, programmers have to deal with protocol relationships "manually". Second, protocol layers can only be assembled through subclassing, and it is thus difficult if not impossible to compose protocol at runtime: in several programming languages, e.g., C++, classes are only compile-time entities. Third, we still have to manage concurrent protocol executions within the same protocol object, while this problem is handled nicely as soon as algorithms are manipulated as objects.

Toolbox Approach

Another possible approach to the reuse of protocol implementations is to provide programmers with a toolbox containing reusable components and associate them with design patterns. Both ASX [21] and CONDUITS+ [13] frameworks can be seen as such toolboxes. The ASX framework provides collaborating C++ components, also known as wrappers, that help in producing reusable communication infrastructures. These components are designed to perform common communication-related tasks, e.g., event demultiplexing, event handler dispatching, connection establishment, routing, etc. Several design patterns, such as the *Reactor* pattern and the *Acceptor* pattern, act as architectural blueprints that guide programmers in producing reusable and portable code. In CONDUITS+ [13], two kinds of objects are basically offered: *conduits* and *information chunks*, which can be assembled in order to create protocol layers and protocol stacks. Various patterns are also provided to help programmers in building protocols.

However, there is no such thing as protocol object in either of the above frameworks. Since our main intent is to provide programmers with a powerful unifying concept, the *protocol object*, we did not choose a toolbox approach for BAST. Furthermore, ASX does not promote protocol composition, whereas CONDUITS+ does it in a slightly different way than BAST, as we discuss below.

Black-box Framework

CONDUITS+ offers basic elements that helps programmers build protocol layers. The use of design patterns is motivated by the fact that traditional layered architectures do not allow code reuse across layers, which is precisely what CONDUITS+ aims at. Protocols can then be composed with CONDUIT+, at lower-level than BAST, through the assembling of conduits and information chunks, which are elementary blocks used to build protocol layers. In other words, the CONDUIT+ framework does not allow the manipulation of protocol layers as objects, but only the manipulation of *pieces* of protocol layers. Compared to BAST, protocol algorithms are further decomposed in CONDUIT+: conduits and information chunks are finer grain objects than BAST's strategies. Indeed, strategies represent protocol layers, while conduits and information chunks are internal components of protocol layers. CONDUIT+ goes one step further in the process of objectifying protocol algorithms.

This approach makes it easy for CONDUIT+ to be a pure *black-box* framework, while BAST combines features of both *black-box* and *white-box* frameworks[7]. With BAST, we are considering completely getting rid of inheritance but this issue has to be carefully studied, because it would have important consequences on the way BAST can be used by application programmers, i.e., those who have

[6] Ingalls and Borning have shown how reflective facilities of Smalltalk can be applied to extend the language with multiple inheritance [14], so we could have used that technique if we really wanted to.

[7] In a *black-box* framework, reusability is mainly achieved by assembling instances, whereas in a *white-box* framework, it is mainly achieved through inheritance. A black-box framework is easier to use, but harder to design.

very limited skills in fault-tolerant distributed algorithms.

Modeling Communications

Several systems model communications but do not really address reliability issues, e.g., STREAMS [18] and the *x*-Kernel [17]. AVOCA [24] defines the notion of protocol objects, but not in the sense that BAST does; furthermore, it mainly applies to high-performance communication subsystems. Other systems offer reliable distributed communications, either based on groups as elemental addressing facilities, e.g., CONSUL [15], ISIS [1] and HORUS [23], or based on transactions, e.g., ARJUNA [22].

Microprotocols and the *x*-Kernel

The work done by O'Malley and Peterson [16] is the closest to BAST that we could find. They extended the *x*-Kernel with the notion of microprotocol graph, and they described a methodology for organizing network software into a complex graph, where each microprotocol encapsulates a single function. In contrast, conventional ISO and TCP/IP protocol stacks have much simpler protocol graphs, with each layer encapsulating several related protocol functions. They argue that such a fine-grain decomposition allows for better tailoring of communication protocols to application needs; our conclusion concurs with theirs perfectly on that point. In their paper, O'Malley and Peterson mainly apply their approach to RPC communications (with only one very short discussion of what they call a *fault-tolerant multicast*). Compared to BAST, their approach is very close to what we have done and is based on the same basic assumption: composing (micro-)protocols is essential when it comes to customizing complex distributed applications (and fault-tolerance implies such complexity). In their terminology, what we call *problem* π is referred to as *metaprotocol* π.

There are also some important differences, however. They do not provide ready-to-use protocol classes to application programmers who are not skilled at understanding and/or building complex protocol graphs, whereas this is one of the main goals of BAST [7]. Moreover, their approach does not rely on design patterns. Similarly to CONDUIT+, they go one step further in their decomposition of protocol algorithms, by defining the notion of *virtual protocols*. The latter "*are not truly protocols in the traditional sense*" [16, page 131] : virtual protocols are actually used to remove IF-statements and to place them in the microprotocol graph instead. All those differences can be best understood by looking at the background domains of the BAST library and the *x*-Kernel respectively. The latter aims at helping system programmers to customize any communication protocol usually found in modern operating systems, while the former aims at providing ready-to-use protocol classes, in order to help any programmer to build fault-tolerant applications, and at allowing skilled programmers to build news fault-tolerant protocols easily.

Composing Protocol Stacks in HORUS

As far as protocol composition is concerned, the HORUS system enables the building of protocol stacks from existing layers only in a strictly vertical manner. Furthermore, it is based on groups as fundamental addressing *and* communication facility, and provides no framework and/or pattern for building *new* protocols layers. HORUS merely provides a finite set of ready-to-use protocol layers, which can only be composed around the group membership protocol.

With BAST, we have tried to model *any kind* of interaction between distributed objects, not only group communications. This is essential in order to deal with failures in an extensible way, because reliable protocols tend to be much more complex than normal communications. By making protocol objects BAST's basic distributed entities, we can build both the group model and the transaction model [6]. Furthermore, the Strategy pattern provides a powerful scheme for creating new protocols through composition.

5 Concluding Remarks

In this paper, we presented how protocol objects can help in building reliable distributed systems. We focused on how the Strategy pattern allows the limitations of inheritance to be overcome, when trying to compose protocols. As far as we know, BAST is the only environment to provide both a set of ready-to-use protocol objects for building fault-tolerant distributed applications, and a complete framework based on design patterns, for composing new protocols from existing ones. We see it as our contribution to the design of well-structured reliable distributed systems.

Our current prototype of BAST is fully operational and is available for Smalltalk and Java. At the moment, inheritance is still partly involved when composing distributed protocols; although a minor drawback, this does not make protocol composition as flexible as one might expect. This is due to the fact that programmers have to know something about the implementation of the protocol classes they reuse, namely their inheritance relationships. This is not surprising, since inheritance is known to violate encapsulation and to hinder modularity. Future work will consist of trying to decide if getting rid of inheritance, at least as far as protocol composition goes, is a good way to achieve even more flexibility. We are also extending BAST with new protocol classes, supporting frequently used protocols in reliable distributed systems, and optimizing existing protocol classes to improve performance. Further information about BAST can be found at *http://lsewww.epfl.ch/bast*; our public-free implementation is also available there.

References

[1] K. Birman and R. Van Renesse. *Reliable Distributed Computing with the Isis Toolkit*. IEEE Computer Society Press, 1993.

[2] T.D. Chandra and S. Toueg. Unreliable failure detectors for reliable distributed systems. *Journal of the ACM*, 34(1):225–267, March 1996.

[3] M. J. Fischer. The consensus problem in unreliable distributed systems (a brief survey). Technical report, Department of Computer Science, Yale University, June 1983.

[4] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. Design patterns: Abstraction and reuse of object-oriented design. In *European Conference on Object-Oriented Programming Proceedings (ECOOP'93)*, volume 707 of *Lecture Notes in Computer Science*. Springer-Verlag, July 1993.

[5] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns, Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1995.

[6] B. Garbinato, P. Felber, and R. Guerraoui. Protocol classes for designing reliable distributed environments. In *European Conference on Object-Oriented Programming Proceedings (ECOOP'96)*, volume 1098 of *Lecture Notes in Computer Science*, Linz (Autriche), July 1996. Springer Verlag.

[7] B. Garbinato and R. Guerraoui. Flexible protocol composition in BAST. Technical report, Operating System Laboratory (Computer Science Department) of the Swiss Federal Institute of Technology, March 1997.

[8] A.J. Goldberg and A.D. Robson. *SMALLTALK-80: The Language and its Implementation*. Addison Wesley, 1983.

[9] J. Gosling and H. McGilton. The Java language environment: A white paper. Technical report, Sun Microsystems, Inc., October 1995.

[10] R. Guerraoui. Revisiting the relationship between non-blocking atomic commitment and consensus. In J.-M. Hélary and M. Raynal, editors, *Distributed Algorithms - 9th International Workshop on Distributed Algorithms (WDAG'95)*, volume 972 of *Lecture Notes in Computer Science*, pages 87–100. Springer Verlag, September 1995.

[11] R. Guerraoui and A. Schiper. Transaction model vs virtual synchrony model: Bridging the gap. In *Theory and Practice in Distributed Systems*, volume 938 of *Lecture Notes in Computer Science*, pages 121–132. Springer Verlag, 1995.

[12] R. Guerraoui and A. Schiper. Consensus service: A modular approach for building agreement protocols in distributed systems. In *Proceedings of the 26th Annual International Symposium on Fault-Tolerant Computing (FTCS-26)*, pages 168–177. IEEE Computer Society Press, June 1996.

[13] H. Hüni, R. Johnson, and R. Engel. A framework for network protocol software. In *Object-Oriented Programming Systems, Languages and Applications Conference Proceedings (OOPSLA'95)*. ACM Press, 1995. Special Issue of Sigplan Notices.

[14] D.H.H. Ingalls and A.H. Borning. Multiple inheritance in smalltalk-80. In *Proceedings of the National Conference on Artificial Intelligence*, pages 234–237. AAAI, 1982.

[15] S. Mishra, L. Peterson, and R. Schlichting. Experience with modularity in Consul. *Software Practice and Experience*, 23(10):1053–1075, October 1993.

[16] S. W. O'Malley and L. L. Peterson. A Dynamic Network Architecture. *ACM Transactions on Computer Systems*, 10(2):110–143, May 1992.

[17] L. Peterson, N. Hutchinson, S. O'Malley, and M. Abott. Rpc in the $x-$Kernel: Evaluating new design techniques. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP'89)*, pages 91–101, November 1989.

[18] D. Ritchie. A stream input-output system. *Bell Laboratories Technical Journal*, 63(8):1897–1910, 1984.

[19] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorenson. *Object-Oriented Modeling and Design*. Prentice-Hall, 1991.

[20] A. Schiper and R. Guerraoui. Faul-tolerant total order "multicast" with an unreliable failure detector. Technical report, Operating System Laboratory (Computer Science Department) of the Swiss Federal Institute of Technology, November 1995.

[21] D.C. Schmidt. ASX: an object-oriented framework for developing distributed applications. In *Proceedings of the 6^{th} USENIX C++ Technical Conference*. USENIX Association, April 1994.

[22] S.K. Shrivastava, G.N. Dixon, and G.D. Parrington. An overview of the Arjuna distributed programming system. *IEEE Software*, 1991.

[23] R. van Renesse, K. Birman, R. Friedman, M. Hayden, and D. Karr. A framework for protocol composition in Horus. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC'95)*, August 1995.

[24] M. Zitterbart, B. Stiller, and A. Tantawy. A Model for High-Performance Communication Subsystems. *IEEE Journal on Selected Areas in Communication*, 11(4):507–519, May 1993.