Developing Multiagent Systems: The Gaia Methodology

Franco Zambonelli[*]   Nicholas R. Jennings[†]   Michael Wooldridge[‡]

October 3, 2003

[*] Dipartimento di Scienze e Metodi dell'Ingegneria,
Università di Modena e Reggio Emilia,
Via Allegri 13 – 42100 Reggio Emilia, Italy.
`franco.zambonelli@unimo.it`

[†] School of Electronics and Computer Science,
University of Southampton, Southampton SO17 1BJ, UK.
`nrj@ecs.soton.ac.uk`

[‡] Department of Computer Science, University of Liverpool,
Liverpool L69 7ZF, UK.
`M.J.Wooldridge@csc.liv.ac.uk`

Abstract

Systems composed of interacting autonomous agents offer a promising software engineering approach for developing applications in complex domains. However, this *multiagent system* paradigm introduces a number of new abstractions and design/development issues when compared with more traditional approaches to software development. Accordingly, new analysis and design methodologies, as well as new tools, are needed to effectively engineer such systems. Against this background, the contribution of this paper is twofold. First, we synthesize and clarify the key abstractions of agent-based computing as they pertain to agent-oriented software engineering. In particular, we argue that a multiagent system can naturally be viewed and architected as a *computational organization*, and we identify the appropriate organizational abstractions that are central to the analysis and design of such systems. Second, we detail and extend the Gaia methodology for the analysis and design of multiagent systems. Gaia exploits the aforementioned organizational abstractions to provide clear guidelines for the analysis and design of complex and open software systems. Two representative case studies are introduced to exemplify Gaia's concepts and to show its use and effectiveness in different types of multiagent system.

1   Introduction

The characteristics and expectations of software systems have changed dramatically in the past few years, with the result that a range of new software engineering challenges have arisen [54, 63]. Firstly, most software systems are now de facto concurrent and distributed, and are expected to interact with components and exploit services that are dynamically found in the network. Secondly, software systems are becoming "always-on" entities that cannot be stopped, restored, and maintained in the traditional way. Thirdly, and as a natural consequence of the first two characteristics, software systems tend to be open, in that they exist in a dynamic operating environment where new components join and existing components leave on a continuous basis, and where the operating conditions themselves are likely to change in unpredictable ways. These characteristics apply, for example, to the semantic web [5], to grid computing [22] and to pervasive environments [1, 54].

Given this new landscape, we advocate the use of *multiagent systems* (MASs) as a software engineering paradigm for designing and developing complex software systems [58, 27, 57]. In MASs, applications are designed and developed in terms of autonomous software entities (*agents*) that can flexibly achieve their objectives by interacting with one another in terms of high-level protocols and languages. These characteristics are well suited to tackling the emerging complexities for a number of reasons. Firstly, the autonomy of the application components (i.e., the ability for an agent to decide what actions it should take at what time [58]) reflects the decentralized nature of modern distributed systems [54] and can be considered as the natural extension to the notions of modularity and encapsulation for systems that are owned by different stakeholders [43]. Secondly, the flexible way in which agents operate (balancing reactive behaviour in response to the environment, with proactive behaviour towards the achievement of their design objectives [58]) is suited to the dynamic and unpredictable situations in which software is now expected to operate [61]. Finally, the high-level and dynamic nature of multiagent interactions is appropriate to open systems in which the constituent components and their interaction patterns constantly change [20, 45, 44].

As is the case with any new software engineering paradigm, the successful and widespread deployment of complex software systems based on MASs requires not only new models and technologies, but also the identification of an appropriate set of software engineering *abstractions*. These abstractions serve as a reference for system identification and analysis, as well as providing the basis of *methodologies* that enable developers to engineer such systems in a robust, reliable, and repeatable fashion. In the last few years, there have been several attempts to identify appropriate abstractions for MASs, and to develop such software engineering methodologies accordingly. However, most of this work is either tuned to specific systems and agent architectures (e.g., [21, 9]) – thus it lacks generality – or else it is defined as a simple extension of existing object-oriented methodologies (e.g., [26, 4]) – and is thus dependent on abstractions and tools that are unsuitable for modeling agent-based systems.

2

Against this background, a number of proposals have recently attempted to define complete and general methodologies, built around agent-specific abstractions and specifically tailored to the analysis and design of MASs [55, 13, 8, 59]. Although these methodologies adopt different terminologies and provide different abstractions, they all recognize that the process of building MASs is radically different from the process of building more traditional software systems. In particular, they all recognize (to varying extents) the idea that a MAS can be conceived in terms of an *organized society* of individuals in which each agent plays specific *roles* and interacts with other agents according to protocols determined by the roles of the involved agents (cf. the more traditional functional composition view of system architectures [48]). However, in this paper we show that an organization is more than simply a collection of roles (as most methodologies assume), and that in order to effectively build a MAS in organizational terms, further organization-oriented abstractions need to be devised and placed in the context of a methodology.

To this end, this paper advances the state of the art in agent-oriented software engineering in two important ways:

- It articulates the role of agent-based computing as a software engineering paradigm and identifies the set of *organizational abstractions* that are necessary for designing and building systems in complex, open environments.

- It extends the Gaia methodology to exploit these new organizational abstractions. These extensions, while preserving the simplicity of the original Gaia proposal, enable it to be used in the analysis and design of open MASs (whereas previously it could only be used for closed communities of cooperating agents).

The remainder of this paper is organized as follows. Section 2 details the key concepts of agent-based computing as they pertain to agent-oriented software engineering. It also outlines the similarities and differences with object- and component- based methods. Section 3 focuses on the organizational metaphor and describes and motivates the organizational abstractions that are necessary for agent-oriented software engineering. Section 4 details how these abstractions are exploited in Gaia in order to provide a methodology for the analysis and design of MASs. Section 5 discusses related work. Section 6 concludes and outlines some open issues.

2   A Software Engineering Perspective on Agent-Based Computing

This section introduces the key concepts of agents and MASs and outline the main differences with work in object/component-based computing and software architectures. The aim is to present the arguments as to why agent-oriented software engineering is a suitable paradigm for designing and building today's complex systems.

2.1 Basic Concepts

The first key concept is that of an *agent*; here viewed as a software entity exhibiting the following characteristics in pursuit of its design objectives [58]:

- *autonomy*: an agent is not passively subject to a global, external flow of control in its actions. That is, an agent has its own internal thread of execution, typically oriented to the achievement of a specific task, and it decides for itself what actions it should perform at what time.

- *situatedness*: agents perform their actions while situated in a particular environment. The environment may be a computational one (e.g., a Web site) or a physical one (e.g., a manufacturing pipeline), and an agent can sense and effect some portions it.

- *proactivity*: in order to accomplish its design objectives in a dynamic and unpredictable environment the agent may need to act to ensure that its set goals are achieved and that new goals are opportunistically pursued whenever appropriate.

For instance, a software component in charge of filtering emails can be viewed as a (simple) agent [34]. It is autonomous if it is implemented as a threaded application logically detached from the client mail reader and if it is assigned the specific task of analyzing and modifying the content of some user mailboxes. It is proactive in that it can draw its user's attention to specific new mails or to specific situations occurring in its folders. It is situated in that it lives in a world of mailboxes and it can sense changes in the mailboxes and can effect their state.

Agents can be useful as stand-alone entities that are delegated particular tasks on behalf of a user (as in the above example and in goal-driven robots [37]). However, in the majority of cases, agents exist in environments that contain other agents. In these *multiagent systems*, the global behaviour derives from the interaction among the constituent agents. This brings us to the second key concept of agent-based computing, that of *sociality* [58]:

- agents interact (cooperate, coordinate or negotiate) with one another, either to achieve a common objective or because this is necessary for them to achieve their own objectives.

Broadly speaking, it is possible to distinguish between two main classes of MASs:

- *distributed problem solving systems* in which the component agents are explicitly designed to cooperatively achieve a given goal;

- *open systems* in which agents are not co-designed to share a common goal, and have been possibly developed by different people to achieve different objectives. Moreover, the composition of the system can dynamically vary as agents enter and leave the system.

The former class of systems includes applications that are primarily devoted to solving computationally intensive problems by assigning the task of exploring different portions of the search space to different agents, with the goal of expediting the search (e.g., [43]). This class also includes those software systems that are devoted to controlling and directing a single physical process (e.g., [9]). Here, different agents are delegated the task of manipulating a sub-portion of the physical space in which the process takes place. In both cases, however, the system is closed; all agents are known a priori, they are supposed to be inately cooperative to each other and, therefore, they can trust one another during interactions. The latter class includes most distributed systems in which the agents have to make use of services, knowledge, and capabilities found in other agents spread throughout the network (e.g., agents for information retrieval [11], for workflow management in virtual enterprises [44], and for pervasive computing [20]), as well as those systems that involve interactions between agents that represent different stakeholders (e.g., e-commerce agents [19] and web service agents [12]). In these cases, one must take into account the possibility of self-interested (competitive) behaviour in the course of the interactions (preventing agents from inherently trusting each other) and the dynamic arrival of unknown agents.

2.2 Objects, Components, and Agents

The distinction between the contemporary view of objects and agents is becoming less sharp with time. However, let us first consider the "traditional" (historical) object-oriented perspective [7]. An object, per se, is not autonomous, in that its internal activity can be solicited only by service requests coming from an external thread of control. As a consequence, an object is not capable of proactive behavior and is not capable of autonomously deciding what to do in a specific situation (i.e., an object is not capable of action selection). Moreover, in traditional object applications, there is not an explicit conception of "environment": objects either encapsulate resources in terms of internal attributes or perceive the world only in terms of other objects' names/references.

On the other hand, objects and components in today's distributed and concurrent systems are somewhat removed from this canonical definition and they are starting to approach our view of agents, at least in terms of observable behaviour. Objects and components may be active and may integrate internal threads of execution enabling them to perform specific computational tasks. They may also have to serve requests in a concurrent way and have to preserve themselves from unauthorized or unsafe requests for services (in other words, they can select not to do an action). Finally, the co-existence in such systems of both active components/objects and passive objects, provides for a clear distinction between an active computational part of the system and a world of external resources (i.e., of an environment). The fact that objects and components have to access such external resources and often have to deal with unexpected situations in so doing (exception and event handling) may require an explicit modeling of situatedness (such as the explicit identification of context-dependencies in component-based systems). Similar considerations can also apply to mod-

ern pervasive computer-based systems. When such entities enter into a complex interactive scenario (e.g., a wireless-enabled building for a PDA [35], a wireless network of distributed sensors [20] or a network of controllers in an automobile [32]) their observable behaviour, from the viewpoint of the system designer, is that of an autonomous agent.

In addition to the fact that modern object and component systems exhibit characteristics that make them assimilable to agents, agents themselves are often implemented in terms of active objects and components (e.g., Java threads with synchronization, exception handling, and event-handling capabilities). However, we emphasize the above mean neither that agents add nothing to modern objects or that the real advantages of agent-based computing can be appreciated only in the presence of AI techniques. Rather, the fact that objects and agents are converging from a technological viewpoint is evidence of the fact that, from a software engineering viewpoint, agent-based abstractions suit the development of complex software systems. This appears even clearer when shifting the attention from single object/components to complex systems.

2.3 Software Architectures and Multiagent Systems

Traditional object-based computing promotes a perspective of software components as "functional" or "service-oriented" entities that directly influences the way that software systems are architected. Usually, the global design relies on a rather static architecture that derives from the decomposition (and modularisation) of the functionalities and data required by the system to achieve its global goals and on the definition of their inter-dependencies [3, 49, 48]. In particular:

- objects are usually considered as service providers, responsible for specific portions of data and in charge of providing services to other objects (the "contractual" model of software development explicitly promotes this view);

- interactions between objects are usually an expression of inter-dependencies; two objects interact to access services and data that are not available locally;

- everything in a system tends to be modeled in terms of objects, and any distinction between active actors and passive resources is typically neglected [47].

In other words, object-oriented development, while promoting encapsulation of data and functionality and a functional-oriented concept of interactions, tends to neglect modeling and encapsulation of execution control. Some sort of "global control" over the activity of the system is usually assumed (e.g., the presence of a single execution flow or of a limited set of controllable and globally synchronised execution flows). However, assuming and/or enforcing such control may be not feasible in complex systems. Thus, rather than being at risk of losing control, a better solution would be to explicitly *delegate* control over the execution to the system components [63] – as in MASs. In fact:

- Delegating control to autonomous components can be considered as an additional dimension of modularity and encapsulation. When entities can encapsulate control in addition to data and algorithms [43], they can better handle the dynamics of a complex environment (local contingencies can be handled locally by components) and can reduce their inter-dependencies (limiting the explicit transfer of execution activities). This leads to a sharper separation between the component-level (i.e., intra-agent) and system-level (i.e., inter-agent) design dimensions, in that also the control component is no longer global.

- The dynamics and openness of application scenarios can make it impossible to know a priori all potential inter-dependencies between components (e.g., what services are needed at a given point of the execution and with what other components to interact), as a functional-oriented perspective typically requires. Autonomous components delegated of their own control can be enriched with sophisticated social abilities, i.e., the capability to make decisions about the scope and nature of their interactions at run-time and of initiating interactions in a flexible manner (e.g., by looking for and negotiating for service and data provision).

- For complex systems, a clear distinction between the active actors of the systems (autonomous and in charge of their own control) and the passive resources (passive objects without autonomous control) may provide a simplified modeling of the problem. In fact, the software components of an application often have a real-world counterpart that can be either active or passive and that, consequently, is better suited to being modeled in terms of both active entities (agents) and passive ones (environmental resources).

Again, we have to emphasize that modern approaches are beginning to recognize the above limitations. As outlined in Subsection 2.2, traditional object abstractions have been enriched by incorporating novel features – such as internal threads of execution, event-handling, exception handling, and context-dependencies – and are being substituted, in architectural styles, by the higher-level abstraction of self-contained (possibly active) coarse-grained entities (i.e., components). These changes fundamentally alter the way software architectures are built, in that active self-contained components intrinsically introduce multiple loci of control are more naturally considered as repositories of tasks, rather than simply of services. Also, the need to cope with openness and dynamics requires application components to interact in more flexible ways (e.g., by making use of external directory, lookup, and security services).

However, attempting to enrich more conventional approaches with novel features and characteristics to meet the novel needs, in addition to increasing the complexity of the modeling, is also likely to introduce a dangerous mismatch between the abstraction level adopted and the conceptual level at which application problems have to be solved. Put simply, objects and components are too low a level of abstraction for dealing with the complexity of today's software systems, and miss important concepts such as autonomy, task-orientation, situatedness and flexible interactions. For

7

instance, object- and component- based approaches have nothing to say on the subject of designing negotiation algorithms to govern interactions, and do not offer insights into how to maintain a balance between reactive and proactive behaviour in a complex and dynamic situations. This forces applications to be built by adopting a functionally oriented perspective and, in turn, this leads to either rather static software architectures or to the need for complex middleware support to handle the dynamics and flexible re-configuration and to support negotiation for resources and tasks. Neither of these are particularly desirable.

In summary, we believe agent-based computing promotes an abstraction level that is suitable for modern scenarios and that is appropriate for building flexible, highly-modular, and robust systems, whatever the technology adopted to actually build the agents.

3 The Organizational Metaphor

Given the suitability of a modeling approach based on autonomous, situated agents that interact in flexible ways, there is a need to understand which further abstractions inspired by which metaphor complete the agent-oriented mindset.

3.1 Motivations

In recent years, researchers in the area of MASs have proposed a number of different approaches for modeling systems based on different metaphors, none of which can reasonably claim to be general-purpose. For instance: the ant algorithms metaphor [6, 2] has shown to be useful in efficiently solving complex distributed problems such as routing and distributed sorting; physical metaphors [1, 35], focusing on the spontaneous reshaping of a system's structure, may have useful applications in pervasive and mobile computing; societal metaphors have been effectively applied in robotics applications [39, 17] and in the understanding and control of highly-decentralized systems [25, 45].

Our approach focuses on the development of medium to large size systems, possibly dived in open and dynamic environments, and that have to guarantee predictable and reliable behaviors. For these kinds of systems, we believe the most appropriate metaphor is that of a *human organization* [24, 23, 18, 61], in which:

- A software system is conceived as the computational instantiation of a (possibly open) group of interacting and autonomous individuals (agents).

- Each agent can be seen as playing one or more specific roles: it has a well-defined set of responsibilities or sub-goals in the context of the overall system and is responsible for pursuing these autonomously. Such sub-goals may be both altruistic (to contribute to a global application goal) or opportunistic (for an agent to pursue its own interests).

8

- Interactions are no longer merely an expression of inter-dependencies, and are rather seen as a means for an agent to accomplish its role in the system. Therefore, interactions are clearly identified and localized in the definition of the role itself, and they help characterize the overall structure of the organization and the position of the agent in it.

- The evolution of the activities in the organization, deriving from the autonomous execution of agents and from their interactions, determines the achievement of the application goal, whether an a priori identified global goal (as, for example, in a workflow management systems where altruistic agents contribute to the achievement of a specific cooperative project), or a goal related to the satisfaction of individual goals (as, for example, in agent-mediated auctions, whose purpose is to satisfy the needs of buyer and seller agents), or both (as, for example, in network enterprises exploiting market mechanisms to improve efficiency).

The organizational perspective leads to a general architectural characterization of a MAS as depicted in Figure 1. Although some simpler systems can be viewed as a single organization, as soon as the complexity increases, modularity and encapsulation principles suggest dividing the system into different sub-organizations (the dashed ellipses in Figure 1), with a subset of the agents being possibly involved in multiple organizations. In each organization, an agent can play one or more *roles*, to accomplish which agents typically need to *interact* with each other to exchange knowledge and coordinate their activities. These interactions occur according to patterns and protocols dictated by the nature of the role itself (i.e., they are institutionalized by the definition of the role). In addition, the MAS is typically immersed in an *environment* (i.e., an ensemble of resources, represented by the gray ellipse in Figure 1) that the agents may need to interact with to accomplish their role. Interactions with the environment occur via some sorts of *sensors* and *effectors* (i.e., mechanisms enabling agents to perceive and act upon some part of the environment). That portion of the environment that agents can sense and effect (represented by the darker ellipses inside the environment in Figure 1) is determined by the agent's specific role, as well as by its current status.

The organizational metaphor – other than being a natural one for human developers who are continuously immersed in a variety of organizational settings and opening up the possibility of re-using a variety of studies and experiences related to real-world organizations [24, 38] – appears to be appropriate for a wide range of software systems. On the one hand, some systems are concerned with controlling and supporting the activities of some (possibly open) real-world organization (e.g., manufacturing control systems, workflow management and enterprise information systems, and electronic marketplaces). Therefore, an organization-based design may reduce the conceptual distance between the software system and the real-world system it has to support. On the other hand, other software systems, even if they are not associated with any pre-existing real-world organization, may have to deal with problems for which human organizations could act as fruitful source of inspiration, having already shown to produce effective solutions (e.g., resource sharing, task as-
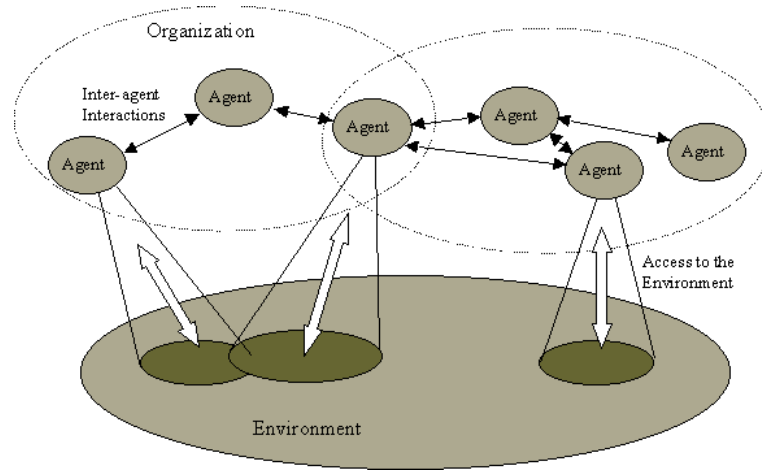
Figure 1: Multiagent Systems as Computational Organizations [27].

signment, and service negotiation). More generally, whenever a software system is complex enough to warrant an agent-based approach and still requires a significant degree of predictability and reliability in all its parts, the organizational metaphor may be the most appropriate one. In fact, by relying on agents playing well-defined roles and interacting according to institutionalized patterns, the organizational metaphor promotes both micro-level (at the agents' level) and macro-level (at the system level) control over the design and understanding of the overall system behavior. Other metaphors (e.g., ant colonies and artificial societies) by focusing on the achievement of an average macro-level behavior of the system, often sacrifice the micro-level aspects (i.e., individual agents matter little or not at all). While this may be acceptable in, say, wide-area file sharing applications and heuristic allocation systems, it is definitely not in manufacturing control systems, enterprise information systems, electronics marketplaces, where each agent is important in its own right.

3.2   Example Applications

To illustrate our arguments, and to show how they map into real-world applications, we will consider two sample problems that will act as running examples throughout this paper.

Manufacturing Pipeline: As an example of the class of distributed problem solving (closed) MASs, we consider a system for controlling a manufacturing process. Specifically, let us consider a manufacturing pipeline in which items are transformed or augmented (e.g., a pipeline in which metal items are painted).

Here, different agents may be devoted to the control of different stages of the pipeline (e.g., one agent is devoted to controlling the paint spraying, another to controlling the heat treatment of the paint, and yet another to controlling the cooling process). In such an organization, the role of each

agent is that of "stage controller", in charge of ensuring that a specific portion of the pipeline works properly (e.g., that the oven maintains a constant temperature and that the cooling system does not cool items too fast). To this end, agents need to sense and effect that portion of the environment which represents the stage of the pipeline of which they are in charge. Since portions of what an agent senses and effects may be sensed and effected by neighboring agents in the pipeline, some sort of indirect interaction, mediated via the environment, is likely to occur between agents. In addition, the agents may need to interact directly with each other to achieve a proper global functioning of the pipeline (for instance, by guaranteeing a uniform flux of items or by guaranteeing that the global flux of items does not exceed the processing capabilities of each of the stages).

Conference Management: As an example of a system raising issues typical of open systems – i.e., agents exhibiting self-interested behavior, not all of whom may be known at deployment time – we consider an agent-based system for supporting the management of an international conference.

Setting up and running a conference is a multi-phase process, involving several individuals and groups. During the submission phase, authors send papers, and are informed that their papers have been received and have been assigned a submission number. In the review phase, the program committee (PC) has to handle the review of the papers: contacting potential referees and asking them to review a number of the papers (possibly by bidding on papers). Eventually, reviews come in and are used to decide about the acceptance or rejection of the submissions. In the final phase, authors need to be notified of these decisions and, in case of acceptance, must be asked to produce a revised version of their papers. The publisher has to collect these final versions and print the proceedings.

The conference management problem naturally leads to a conception of the whole system as a number of different MAS organizations, one for each phase of the process. In each organization, the corresponding MAS can be viewed as being made up of agents being associated to the persons involved in the process (authors, PC chair, PC members, reviewers) to support their work, and representing the active part of the system. The roles played by each agent reflect the ones played by the associated person in the conference organization. This may require agents to interact both directly with each other and indirectly via an environment composed of (passive) papers and review forms[1].

The openness of the conference management system is evidenced in a number of ways. Clearly, authors submitting a paper are unknown at deployment time, and so are the total number of author agents being involved in the process. However, this appears as a not very problematic issue, in that author agents (generally) have simply to send a paper to the PC chair, and do not need to strongly

---

[1]Such modeling differs considerably from traditional, object-based modeling of the same problem (see e.g. [47]), where the same application problem is typically modeled without making any distinction between the active entities and environmental resources (despite their being very different in nature) and where interactions are used to express functional inter-dependencies rather than the organizational relationships that actually cause them.

interact with other agents in the system. More problematic is the fact that, since agents are associated to different stakeholders with competing interests, unpredictable strategic and opportunistic behaviors can emerge in the application. The trivial case being that of a reviewer agent acting on behalf of a person which is also an author, and trying (e.g., by bidding for papers) to review its own paper or possibly competing papers. In addition, we also emphasize that a reviewer (or a PC member) may decide to exploit its own personal agent – rather than the agents made available by the conference management system – to enter the organization and interact with the other agents in it. In this case, even if all the reviewers are known in advance (and we know this is not the case in large conferences) the agents that will interact in the system may not be.

3.3 Organizational Abstractions

The characterization of a MAS of Figure 1 highlights the most basic abstractions that characterize a computational organization and that can be appropriately exploited in the analysis and design phases: the *environment* in which the MAS is immersed; the *roles* to be played by the different agents in the organization; and the *interactions* between these roles. In addition, we have identified two further abstractions that are often implicitly integrated into the above ones and that, we believe, need to be considered in their own right: *organizational rules* and *organizational structures*.

3.3.1 The Environment

A MAS is always situated in some environment and we believe this should be considered as a primary abstraction during the analysis and design phases. Generally speaking, identifying and modeling the environment involves determining all the entities and resources that the MAS can exploit, control or consume when it is working towards the achievement of the organizational goal. In some cases, the environment will be a physical one, including such things as the temperature in a room, the status of a washing machine or the average speed of cars on a motorway. This is the case in the manufacturing pipeline, where the agents are intended to control the correct processing and the global flux of the items in the pipeline. In other cases, the environment will be a virtual one, including such things as enterprise information systems, web services, and database management systems. This is the case of the conference management system, where agents execute in an environment populated by papers, review forms and digital libraries. Whatever the case, an explicit modeling of the environment is very important: not taking it into account (as, for example, in the previous version of Gaia methodology [59]) may complicate the overall design and may introduce mismatches between the MAS design and its actual operation. For instance, in the conference management example, the lack of environmental modeling would imply considering specific application agents as repositories of papers and review forms and as being in charge of explicitly transferring them to other agents. This is a more complex solution than that of explicitly modeling an environ-

mental data repository, and conceptually farther from the likely operating environment (a world of Web sites from which agents can access papers and review forms).

For both physical and computational environments, the following issues come to the fore when considering an explicit environmental modeling phase:

- *What are the environmental resources that agents can effectively sense and effect?* Constraints may exist with respect to the possibility of accessing and manipulating resources, and the environment model should therefore distinguish between the existence and the accessibility of a resource. In the manufacturing pipeline, it may be the case that no sensors exist to effectively evaluate the quality parameters. In conference management, it may be the case that a reviewer agent cannot pay to access a particular digital library (as needed to compare a submitted paper to an already published one). Also, since the environment may be characterised by its own dynamic, it may be the case that some of the resources are ephemeral in nature or that are only intermittently available. These dynamics may need to be identified, modeled, and evaluated against technological capabilities to cope with them.

- *How should the agent perceive the environment?* In other words, what representation of the environment is appropriate in a given situation? Such a choice is naturally dependent on the available technologies and on the pre-existing scenarios. In the manufacturing pipeline, the choice may depend on whether sensors are passive or active, leading to a characterization of the environment in terms of either a data world or a service world. In conference management, such a choice may depend on the characteristics of the available services providing access to papers and review forms.

- *What in the existing scenario should be characterized as part of the environment?* The existence of active entities with which the agents in the MAS will have to interact (e.g., computer-based sensors in the manufacturing pipeline or active databases in a Web scenario), means a decision has to be taken about what should be viewed as an agent and what should be viewed in terms of dynamic environmental resources. In other words, the distinction between the agent and the environment is not always clear cut. It is something that may require an accurate analysis and may ultimately depend on the problem's characteristics.

Summarizing, the environment of a MAS should not be implicitly assumed: its characteristics must be identified, modeled, and possibly shaped to meet application-specific purposes.

3.3.2 Roles and Interactions

The role of an agent defines what it is expected to do in the organization, both in concert with other agents and in respect of the organization itself. Often, an agent's role is simply defined in terms of

the specific task that it has to accomplish in the context of the overall organization. However, our notion of a role is much more precise; it gives an agent a well-defined position in the organization, with an associated set of expected behaviours [59, 21, 18]. Organizational role models precisely describe all the roles that constitute the computational organization. They do this in terms of their functionalities, activities, and responsibilities, as well as in terms of their interaction protocols and patterns. Organizational interaction models describe the protocols that govern the interactions between the roles. Moreover, the interaction model describes the characteristics and dynamics of each protocol (e.g., when, how, and by whom a protocol has to be executed).

Most approaches to MAS modeling [33, 55], (including the first version of Gaia [59]), consider role and interaction models as the sole organizational abstractions upon which to base the entire development process. Consequently, the analysis and design process starts directly with the identification and definition of role and interaction models. However, although role and interaction models can be useful to fully *describe* an existing organization, they are of limited value in *building* an organization. In fact, before the design process can define an actual organization (to be possibly described in terms of role and interaction models), there is a need to identify *how* the organization is expected to work and *which* kind of organization (among several possible ones) best fits the requirements identified in the analysis phase. This observation – elaborated below – motivates our introduction of the notions of organizational rules and organizational structures.

3.3.3 Organizational Rules

In the requirements capture phase of MAS development, it is certainly possible to identify the basic skills (functionalities and competences) required by the organization, as well as the basic interactions that are required for the exploitation of these skills. In the case of the manufacturing pipeline, it is easy to identify the need for agents to act as "stage controllers"; in the conference management case study, it is easy to recognize the need for agents to act as "authors", "reviewers", "PC Chair", and so on.

Although such an analysis can lead to the identification of a *preliminary* version of the role and interaction models, this identification cannot and *should not* be used to produce complete models. In fact, before being able to fully characterize the organization, the analysis of a MAS should identify the constraints that the actual organization, once defined, will have to respect. Typically, such constraints: *(i)* spread horizontally over all the roles and protocols (or, which is the same in this context, over the identified preliminary roles and protocols), or *(ii)* express relations and constraints between roles, protocols, or between roles and protocols [19]. Although, in an actual organization, such constraints are likely to be somehow enacted by some agents playing some roles and interacting somehow, they can hardly be expressed in terms of individual roles or individual interaction protocols (in the same way that social conventions and company directives horizontally influence our social life and our work, but cannot be associated with any specific actor). Thus, the explicit

14

identification of such constraints – captured in our concept of *organizational rules* – is very important for the correct understanding of the characteristics that the organization-to-be must express and for the subsequent definition of the system structure by the designer.

The explicit identification of organizational rules is also important in the context of open systems. With the arrival of new, previously unknown, and possibly self-interested agents, the overall organization must be able to enforce its internal coherency despite the dynamic and untrustworthy environment. The identification of global organizational rules allows the system designer to explicitly define: *(i)* whether and when to allow new agents to enter the organization, and, once accepted, what their position should be; and *(ii)* which behaviours should be considered as a legitimate expression of self-interest, and which among them must be prevented by the organization. In this context, organizational rules may also drive the designer towards the definition of the specific organization structure that most eases the enforcement of the organizational rules and, for instance, can facilitate the prevention of undesirable behaviour on the part of the unknown agents.

In the manufacturing pipeline example, all the different stages must maintain the same speed of flow of items in the pipeline. This requirement is most simply expressed in terms of a global organizational rule, rather than replicating it as a requirement for each and every role in the organization (as, for example, the previous version of Gaia and any methodology not making explicit use of organizational rules would have required). In the conference management system, there are a number of rules that drive the proper implementation of the organization. As notable examples: an agent should be prevented from playing both the role of author and reviewer of the same paper; PC members should not be in charge of collecting the reviews for their own papers. Neither of these constraints can easily be expressed in terms of properties or responsibilities associated to single roles or protocols and, if so, they would notably increase the complexity of roles and protocol description. Instead, they represent global organizational rules.

3.3.4 Organizational Structures

A role model describes all the roles of an organization and their positions in that organization. Therefore, a role model also implicitly defines the *topology* of the interaction patterns and the *control regime* of the organization's activities. That is, it implicitly defines the overall architecture of the MAS organization (i.e., its *organizational structure*). For example, a role model describing an organization in terms of a "master role" and "slave roles" – where the former is in charge of assigning work to the latter and of load balancing their activities – implicitly defines an organizational structure based on a hierarchical topology and on a load partitioning control regime. Other exemplar organizational structures include collectives of peers, multi-level and multi-divisional hierarchies [23], as well as more dynamic structures deriving from market-oriented models [31]. All these organizations can be modeled in terms of a role model.

However, while the role model may define the organizational structure in an implicit way, the

15

structure of a MAS is more appropriately derived from the explicit choice of an appropriate organizational structure, and organizational structures should be viewed as first-class abstractions in their own right. This argument (which is conventional in architecture-centered software design [3, 49]) calls for a specific design choice not to be (generally) anticipated to the analysis phase, but rather should exploit information collected during analysis. In the specific context of MAS development, the argument is motivated by several considerations:

- Although the organizational structure of a MAS may be directly inspired by the structure of the real-world system that the MAS must support, automate, or monitor (consider, for example, enterprise information systems and workflow management systems), this should not automatically imply that the organization of the software system should always mimic that of the real world system. This is so for several reasons. First, the real world organization may not be necessarily so well-structured. Second, the issues that may have driven a human organization towards the adoption of a particular structure may not necessarily apply to the agent organization. Third, the mere presence of software may introduce changes in the real-world organization. Of course, it may be the case that some specific sub-structures in a real-world organization are to be necessarily preserved at the MAS level and, thus, come pre-defined from the analysis phase. However, this should not be considered the general case and should not prevent developers from explicitly addressing the organizational structure issue.

- Starting from the organizational structure may prevent optimizing the overall efficiency of the organization and may prevent subsequent optimizations and changes. This consideration assumes a particular emphasis in dynamic applications scenarios (e.g., virtual enterprises [44], global [2] and pervasive computing [20, 54]) where a system may need to frequently adapt its organizational structure to the prevailing situation, possibly at run-time and in an unsupervised way [30]. Although this paper does not explicitly deal with dynamic and unsupervised re-organizations, the methodological approach we propose (by making the definition of the organizational structure an explicit design decision) facilitates off-line re-organization and paves the way for supporting dynamic on-line re-organization.

- The organization, once defined, has to respect its organizational rules. Starting from a pre-defined organizational structure – by assuming to know in advance what it should be or by committing a priori to a given organizational structure – can make it difficult to have the organizational rules respected and enforced. It is more natural for the choice of the organizational structure to follow from the identification of the organizational rules.

In the manufacturing pipeline, perhaps the most natural choice is to have an organizational structure in which all of the stages in the pipeline form a collective of peers. For instance, with reference
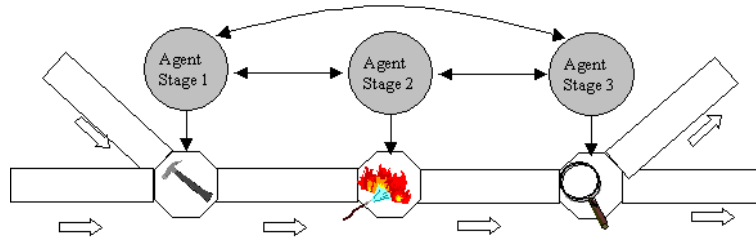
16

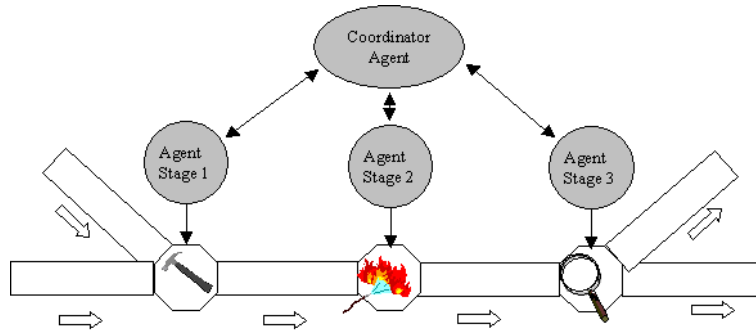Figure 2: Manufacturing pipeline: collective of peers organization

Figure 3: Manufacturing pipeline: hierarchical organization

to Figure 2, the various stages of the pipeline are controlled by agents specifically devoted to controlling one stage, and each of these agents may directly interact with all the other stages to agree on issues requiring global coordination (e.g., regulating the flux of items in the pipeline). In other words, the topology of the interactions is a fully connected one and is subject to a fully distributed cooperative control regime. However, this is neither the only possible choice, nor necessarily the best one. For instance, because of the real-time nature of the pipeline control problem, it may be the case that some problem requiring complex global coordination between all the agents cannot be solved in time because of the high coordination costs associated with peer-to-peer interactions. In such cases, the designer can adopt a different organizational structure. For example, as sketched in Figure 3, a global coordinator agent can be introduced to control and mediate the interactions for all the other agents. This, in turn, leads to a hierarchical organization.

In the conference management, the overall structure of the MAS can generally be derived directly from the structure the conference officials have explicitly decided to adopt. A small conference usually relies solely on the PC members for the review process, with the PC chair acting as a global coordinator in a single-level hierarchy to control the reviewing work of the PC members (see Figure 4). The higher workload in a large conference usually requires a different approach (see Figure 5). For instance, the PC chair may partition the papers among the PC members, and the PC members may be in charge of finding reviewers for the papers in their partitions (either by asking a reviewer to review a specific paper or, alternatively, by calling for bids on papers in a sort
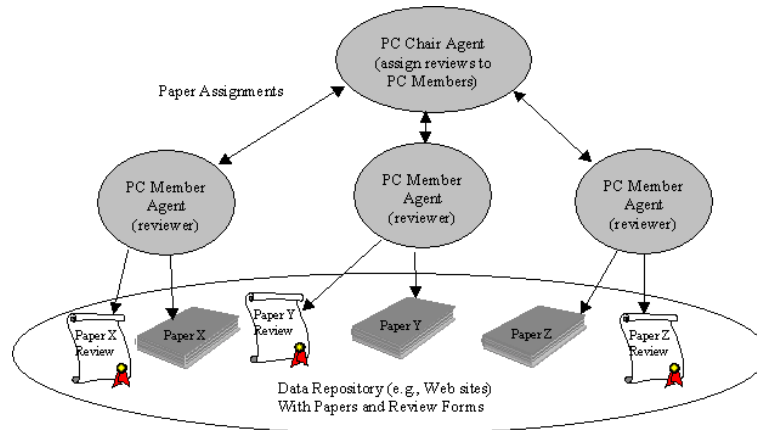
Figure 4: Conference management: single level hierarchy

of contract-net protocol [52]). Also, PC members may negotiate with one another for re-allocating papers to different partitions. In other words, the organizational structure can be composed by a hierarchy at the higher level (ruling interactions between the PC Chair and the PC Members), and by a "market" organization at the lower level (to ensure a proper assignment of papers to reviewers and to enable direct negotiations between the collective of PC member agents). However, the development of a MAS for a conference must take into account the fact that the conference may change its dimensions (and its structure) from year to year and that the support of an agent-based system (reducing the work of the conference officials) may suggest a different organizational choice than that originally planned. Thus, if the analysis describes the system's requirements without committing to a specific organizational structure (as we suggest), the designer can reuse the analysis work to produce a new design according to the conference's new needs.

As an additional note related to organizational structures, we believe that despite the huge number of structures that can possibly be conceived, a (comparatively) small subset of these structures are likely to be used most of the time. This opens up significant opportunities both for re-use and for the exploitation of catalogues of agent-oriented *organizational patterns* – including use-cases reporting on efficiency, robustness, degree of openness, and ease of enactment of organizational structures – to support designers in choosing the most appropriate organizational structure for their problem. For instance, in the manufacturing example, the collective organization expresses a pattern that is likely to re-appear in many applications. The same can also be said of the hierarchical structure. In the conference management example, the various organizational structures that conferences of different sizes tend to adopt are also all fairly typical: single hierarchies, composite ones, contract nets, etc. Currently, in the area of agent-oriented software engineering, most pattern-related work focuses on detailed design patterns [53, 29]. In the near future, we expect more research and studies to be carried on in the area of architectural, organization-oriented patterns, to extend the specific studies already performed in this area regard [23, 31] and to adapt those studies performed in the
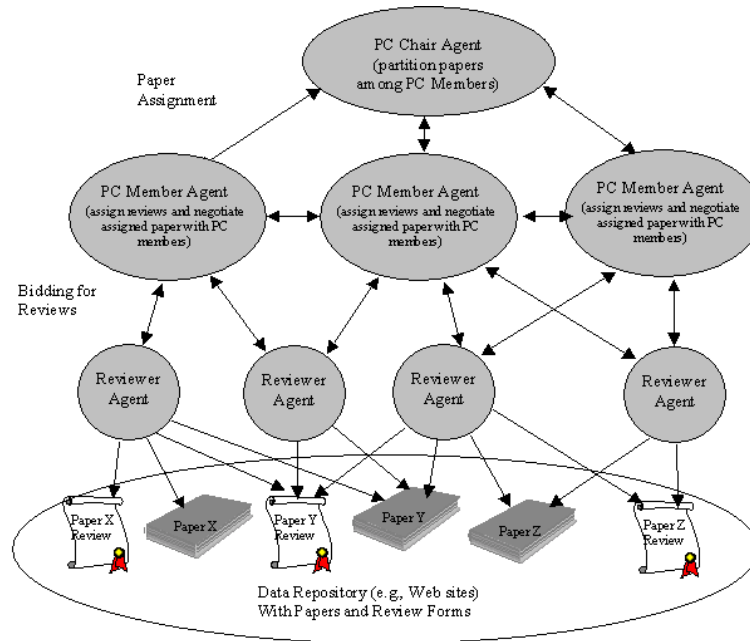
18

Figure 5: Conference management: multi-level hierarchy

area of organization management [38, 24] for exploitation in agent-oriented methodologies.

4 The Gaia Methodology

Having introduced and defined the various organizational abstractions that we believe are necessary for analyzing and designing MASs, the next step is to fashion them into a *design process*. That is, to produce an ordered sequence of steps, an identifiable set of models, and an indication of the interrelationships between the models, showing how and when to exploit which models and abstractions in the development of a MAS. The design process that we propose uses our previous work on the Gaia methodology [59] as a point of departure. The new, extended version of Gaia (simply called Gaia hereafter) exploits the new organizational abstractions we identified in Section 3 and significantly extends the range of applications to which Gaia can be applied.

Before going into the details of Gaia, we first provide an overview of its key stages and models (see figure 6). The Gaia process starts with the analysis phase, whose aim is to collect and organize the specification which is the basis for the design of the computational organization. This includes the identification of:

- The goals of the organizations that constitute the overall system and their expected global behaviour. This involves identifying how to fruitfully decompose the global organization into loosely-coupled sub-organizations.
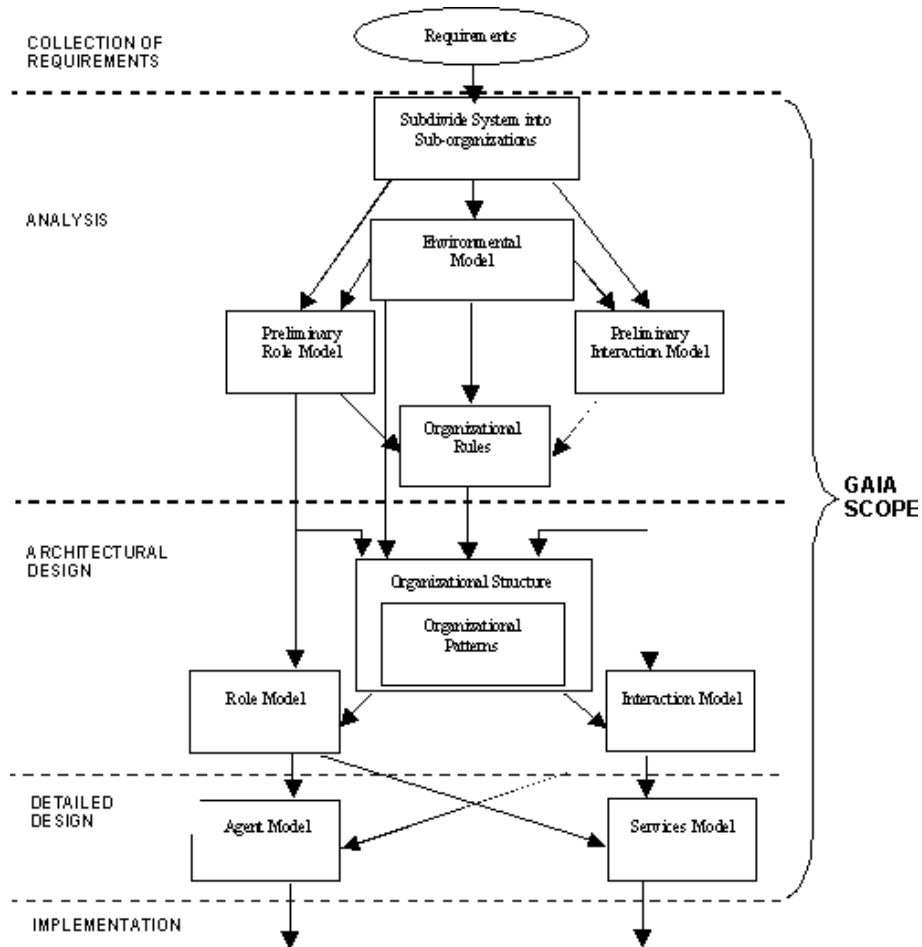
19

Figure 6: Models of the Gaia Methodology and their Relations in the Gaia Process

- The environmental model; intended as an abstract, computational representation of the environment in which the MAS will be situated.

- The preliminary roles model; identifying the basic skills required by the organization. This preliminary model contains only those roles, possibly not completely defined, that can be identified without committing to the imposition of a specific organizational structure. Also, the notion of roles, at this stage, is abstract from any mapping into agents.

- The preliminary interaction model; identifying the basic interactions required to accomplish the preliminary roles. Again, this model must abstract away from the organizational structure and can be left incomplete.

- The rules that the organization should respect and enforce in its global behaviour. Such rules express constraints on the execution activities of roles and protocols and are of primary

20

importance in promoting efficiency in design and in identifying how the developing MAS can support openness and self-interested behaviour.

The output of the analysis phase – consisting of an environmental model, a preliminary roles model, a preliminary interactions model, and a set of organizational rules – is exploited by the design phase, which can be logically decomposed into an architectural design phase and a detailed design phase. The architectural design phase includes:

- The definition of the system's organizational structure in terms of its topology and control regime. This activity, that could also exploit of catalogues organizational patterns, involves considering: *(i)* the organizational efficiency, *(ii)* the real-world organization (if any) in which the MAS is situated, and *(iii)* the need to enforce the organizational rules.

- The completion of the preliminary role and interaction models. This is based upon the adopted organizational structure and involves separating – whenever possible – the organizational-independent aspects (detected from the analysis phase) and the organizational-dependent ones (derived from the adoption of a specific organizational structure). This demarcation promotes a design-for-change perspective by separating the structure of the system (derived from a contingent choice) from its goals (derived from a general characterization).

Once the overall architecture of the system is identified together with its completed roles and interactions model, the detailed design phase can begin. This covers:

- The definition of the agent model. This identifies the *agent classes* that will make up the system and the *agent instances* that will be instantiated from these classes. There may be a one-to-one correspondence between roles and agent types, although a number of closely related roles can be mapped into in the same agent class for the purposes of convenience or efficiency.

- The definition of the services model. This identifies the main services – intended as coherent blocks of activity in which agents will engage – that are required to realize the agent's roles, and their properties.

Before detailing the analysis (Subsection 4.1), the architectural design phase (Subsection 4.2) and the detailed design phase (Subsection 4.3), it is important to explicitly state the scope and limitations of our methodology:

- Gaia does not directly deal with particular modeling techniques. It proposes, but does not commit to, specific techniques for modeling (e.g., roles, environment, interactions). At this time, and given the large amount of ongoing research work devoted to defining suitable (and possibly standardized) notation techniques for agent systems (as detailed in Section 5), we believe such a commitment would be premature.

- Gaia does not directly deal with implementation issues. The outcome of the Gaia process is a detailed but technology-neutral specification that should be easily implemented using an appropriate agent-programming framework (e.g., a FIPA-compliant agent system) or by using a modern object/component-based framework (i.e., one supporting distributed and concurrent objects). Of course, we are aware that specific technology platforms may introduce constraints over design decisions (a typical example relates to environmental modeling: if the environmental resources are of a passive nature, one cannot rely on having agents perceive them in terms of events). However, generally speaking, these situations should be known at the time of requirements capture.

- Gaia does not explicitly deal with the activities of the requirements capturing and modeling, and specifically of early requirements engineering [40]. We are aware of the importance of such activities and of the ongoing work in this area (see section 5). In particular, we believe that Gaia could fit and be easily integrated with modern goal-oriented approaches to requirements engineering [14], whose abstractions closely match those of agent-oriented computing. However, the investigation of such an issue is beyond the scope of this paper.

4.1 The Analysis Phase

The main goal of the analysis phase is to organize the collected specifications and requirements for the system-to-be into an environmental model, preliminary role and interaction models, and a set of organizational rules, for each of the (sub-organizations) composing the overall system.

The basic assumption in Gaia is that the analysis phase can rely on the output produced by an early requirements engineering phase, devoted to analyzing the characteristics to be exhibited and the goals to be achieved by the system-to-be, as they emerge from the needs of the stakeholders and from the specific operational environment. The increasing acceptance of goal-oriented approaches to early requirements engineering [40] – modeling the specifications of systems in terms of the actors' involved, their roles, and their goals – is well suited to the organizational abstractions exploited by Gaia. Moreover, it may facilitate the Gaia analysis (which can accordingly be considered as a form of late requirements engineering phase), as well as the subsequent design phases.

4.1.1 The Organizations

The first phase in Gaia analysis is concerned with determining whether multiple organizations have to co-exist in the system and become autonomous interacting MASs. Identifying such organizations is reasonably easy if *(i)* the system specification already identifies them or *(ii)* the system mimics the structure of the real world, and this involves multiple, interacting organizations. However, even if neither of these conditions hold, modularity concerns may suggest considering the system in terms of multiple organizations, for the sake of splitting the global complexity of a problem into a set

of smaller more manageable components [51]. Generally speaking, such sub-organizations can be found when there are portions of the overall system that *(i)* exhibit a behaviour specifically oriented towards the achievement of a given sub-goal, *(ii)* interact loosely with other portions of the system, or *(iii)* require competences that are not needed in other parts of the system.

Turning to our running examples. The agents controlling the manufacturing pipeline are likely to be conceived as belonging to the same organization: *(i)* they share the same overall goal (making the pipeline run efficiently), *(ii)* they sometimes need to interact intensively with each other, while they interact less frequently with those agents concerned with controlling other parts of the factory (e.g., the electric plant), and *(iii)* the required agent skills are very closely related to that section of the manufacturing process. With regard to the conference management example, three sub-organizations can be clearly identified. Firstly, the organization responsible for the submission process is in charge of distributing the call for papers, collecting the submitted papers, and possibly doing some preliminary review/control of the papers. Secondly, the organization responsible for the review process. This is in charge of assigning submitted papers to reviewers, of collecting and ranking the reviews, and defining the technical programme. Thirdly, the organization responsible for the publication of the proceedings. In all cases, there is a clear sub-goal to be pursued in each organization, the interactions between organizations are loose and scheduled at specific intervals of time, and there are entities that participate in some organizations and not in others.

4.1.2   The Environmental Model

It is difficult to provide general modeling abstractions and general modeling techniques because (as already stated in Section 3.1) the environments for different applications can be very different in nature and also becase they are somehow related to the underlying technology. To develop a reasonably general approach (without the ambition for it to be universal), we suggest treating the environment in terms of *abstract computational resources*, such as variables or tuples, made available to the agents for sensing (e.g., reading their values), for effecting (e.g., changing their values) or for consuming (e.g., extracting them from the environment).

Following such identification, the environmental model (in its simplest form) can be viewed as a list of resources; each associated with a symbolic name, characterized by the type of actions that the agents can perform on it, and possibly associated with additional textual comments and descriptions. A simple way to represent these is as follows (inspired by the FUSION notation for operation schemata [16, pp26–31]):

$$
\begin{array}{lll}
\texttt{reads} & \textit{Var}1 & \textit{// readable resource of the environment} \\
& \textit{Var}2 & \textit{// another readable resource} \\
\texttt{changes} & \textit{Var}3 & \textit{// a variable that can be also changed by the agent}
\end{array}
$$

Returning to our running examples. In the manufacturing pipeline, and assuming that the MAS is being deployed with the sole goal of ensuring the correct flow of items in the pipeline, the resources of interest are the number of items being produced and made flow in each of the pipeline's n stages (other resources such as temperature, humidity, position, may be relevant for other specific activities of the stages). Thus, if appropriate sensors and actuators are available to detect and control, respectively, such flux, the environmental model reduces to:

$$\texttt{changes} \quad \textit{flux}[i], i = 1, n \quad \textit{// number of items flowing in each stage of the pipeline}$$

In the conference management, the resources handled by the agents are papers and review forms. In a minimal (exemplifying) representation, the environmental model can be as follows:

$$\texttt{reads} \quad \textit{papers}[i], i = 1, \textit{total submitted} \qquad \textit{// all papers submitted for review}$$
$$\texttt{changes} \quad \textit{reviews}[i][j], i = 1, \textit{total submitted}; j = 1, \textit{number of reviewers} \quad \textit{// reviews for the submitted papers}$$

Clearly, in realistic development scenarios, the analyst would choose to provide a more detailed and structured view of environmental resources. For instance, a paper would be represented by a data structure including information such as authors, title, keywords, and so on. Also, it is worth pointing out that more specific modeling techniques may be better adopted depending on specifications and technological constraints. For instance, the development of Web-based information based on XML documents (as may be the case in a conference management system) might profit from an environmental model by preserving the already available XML data representation and simply enriching it with annotations related to how the MAS will access these structures.

In any case, there are other specific issues related to the modeling of the environmental resources that may be required to enrich/complement our proposed basic notation. In particular:

- Often, due to either the distributed nature of the environment or to the logical/physical relationships existing between its resources, a graphical (possibly annotated) scheme may be of help to represent such relationships and to identify how and from where a resource can be accessed. In the manufacturing pipeline, the various stages are intrinsically distributed in space, and it may be the case that the data related to the flux of an item in a particular stage can only be accessed by the stage itself. In this case, a graphical representation as per Figure 7 facilitates the capturing and understanding of the environmental characteristics. A similar representation can be used for the conference management environment where papers and review forms may be grouped either according to some logical relationships (e.g., the type of submission) or their network distribution (e.g., when a multiple track conference invites
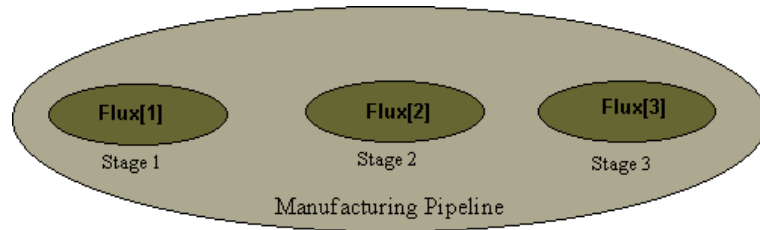
Figure 7: Graphical representation of the manufacturing pipeline environment

submissions directly to the Web sites of the track chairs).

- The environment in which a MAS is immersed is typically characterized by its own dynamics. On the one hand, the content of the environmental resources may change over time according to patterns that may be of some relevance and may require some explicit modeling (for the sake of, for example, analytical modeling). These cases can generally be dealt with via annotations to the basic environmental model. On the other hand, some of the resources could be intermittently available, of an ephemeral nature, or simply be a priori unknown. These cases may also require enriching the basic access model to resources to take into account such uncertainties. For instance, an associative access model as per Linda tuple spaces may well suit this purpose [11], as proved by its increasing adoption in open distributed systems [12].

An important issue that should be borne in mind when modeling the environment is the fact that the operational environment of a MAS may include active components (i.e., services and computer-based systems) with which agents in the MAS have to interact. This introduces the problem of how to deal with such entities: should they be modeled as resources of the environment, or should they instead be "agentified"? (i.e., modeled in terms of additional agents in the MAS) Some general guidelines can be provided with respect to this issue:

- When the role of these active components is simply that of a data provider (consider, for example, a Web server or a DBMS mediating access to a dataset, or a simple computer-based sensor), it is better to abstract away from their presence and to model them in terms of resources. The rationale for this is that their presence influences only the mechanisms by which agents retrieve resources (i.e., obtaining the data by requesting a service rather than by performing a sensing operation), not the nature of the resources themselves or the internal activities of the agents. Similar considerations may apply for simple components capable of event-notification services (i.e., upon change of a resource value).

- If the environment contains components and services that are capable of performing complex operations (e.g., active databases, active control systems, humans in-the-loop) then their effects on the agents' perception of the environment can make it hard to model them as a simple

resource repository with identifiable patterns of dynamic change to be sensed by agents (or to interact with based on event-notification mechanisms). In such cases, these components should not be treated as part of the environment but, instead, they should be agentified.

4.1.3 The Preliminary Role Model

The analysis phase is not intended to design the actual organization of the MAS (this is the purpose of the subsequent architectural design phase). However, even without knowing what the structure of the organization will be, it is possible, in most cases, to identify some characteristics of the system that are likely to remain the same independently of the actual organizational structure. In particular, this equates to identifying the "basic skills" that are required by the organization to achieve its goals, as well as the basic interactions that are required for the exploitation of these skills. Such identification activities may even be facilitated if a goal-oriented early requirements analysis has already modeled the characteristics of the system in terms of actors involved and their goals.

Given the identification of the basic skills and of their basic interaction needs, respectively, the analysis phase can provide a *preliminary definition* of the organization's roles and protocols. However, this definition cannot be completed at this stage. In fact, the basic skills (or *preliminary roles*) can only truly become organizational roles when it is known how and with which other roles they will interact. This, in turn, requires the definition of the global organizational structure. Analogously, the basic interaction needs (or *preliminary protocols*) can only be fully defined as organizational protocols when the adoption of an organizational structure clarifies which roles these protocols will involve, when their execution will be triggered, and by whom. In addition, the design phase is likely to introduce additional roles and protocols, directly derived from the adopted organizational structure and, therefore, not identified in the analysis phase.

As anticipated in Section 3.3.3, it may also be the case that the need to adopt specific organizational sub-structures in a system derives from specific requirements, in which case at least some of the roles and protocols may assume an already complete definition even at this stage.

In the manufacturing pipeline, the preliminary roles (PR) that can be devised are those associated with each of the stages in the pipeline:

PR = STAGE[1], STAGE[2], . . . , STAGE[N].

In addition, depending on the global structure of the manufacturing process, it is possible to devise other preliminary roles. For example, when the flux of items in the pipeline require more power than is currently supplied, an additional role may have to be introduced whose task is to coordinate with the electric plant (better, the agent organization of the electric plant) and negotiate further supplies of energy. However, unless the design has identified whether the overall system should be structured as an organization of peers (Figure 2) or a hierarchy (Figure 3) the role specification cannot be completed, nor is it possible to identify all the roles involved in the organization. Analogously, it may be clear from the analysis phase that a protocol should be defined for enabling

one stage to ask for a slowing down in the flux of items in the pipeline. However, in this case, it is not certain which other entity this protocol should involve (the other stages in the pipeline or a global controller, depending on the chosen organizational structure).

In the conference management case, and in particular in the organization of the review process, it is comparatively easy to devise a number of preliminary roles that will be played whatever the organizational structure. For example, the roles associated with selecting reviewers and assigning papers to them (REVIEWCATCHER), the role of reviewer in charge of filling review forms for assigned papers (REVIEWER), the role in charge of collecting and ranking the reviews (REVIEWCOLLECTOR) and the role of finalizing the technical programme (DOPROGRAM). Depending on the actual organizational structures, different actors (e.g., the PC Chair, the PC Members or external reviewers) will be called to play such roles. However, the actual organizational structure, defined in the architectural design phase, may also require additional roles to be introduced. For example, a big conference may opt to subdivide the submitted papers among a set of co-chairs according to their competences. In this case, an additional role is required, for partitioning the submitted papers according to the defined criterion. This is likely to influence the definition of the protocols and to require the definition of further protocols (for example, protocols to achieve some form of load balancing on the partitions).

To represent (preliminary) roles, Gaia adopts an abstract, semi-formal, description to express their capabilities and expected behaviors. These are represented by two main attribute classes, respectively: *(i)* permissions and *(ii)* responsibilities.

Permissions: These attributes are mainly aimed at: *(i)* identifying the resources that can legitimately be used to carry out the role – intuitively, they say what *can* be spent while carrying out the role; and *(ii)* stating the resource limits within which the role must operate – intuitively, they say what *can't* be spent while carrying out the role. In general, permissions relate agent roles to the environment in which they are situated: in order to carry out a role, an agent will typically have to access environmental resources and possibly change or consume them. However, they can also be used to represent *knowledge* the agent can possibly have or have received from communications with other roles. To represent permissions, Gaia makes use of the same notation already used for representing the environmental resources. However, the attributes associated with resources no longer represent what can be done with such resources (i.e., reading, writing, or consuming) from the environmental perspective, but what agents playing the role must be allowed to do to accomplish the role and what they must not be allowed to do.

In the manufacturing pipeline example, an agent in charge of controlling a specific stage of the pipeline may need to sense (read) the abstract resource representing the flux of items entering its region of responsibility. However, since the previous stage is controlled by a different agent, it cannot control (change) it. Nevertheless, the role requires an agent to sense and effect the flux of items in its stage, representing the input flux to the next stage. This reduces to the following

representation of the permissions for a generic role STAGE[i]:

$$\begin{array}{lll} \texttt{reads} & \textit{flux}[i-1] & \textit{// flux from previous stage} \\ \texttt{changes} & \textit{flux}[i] & \textit{// flux to next stage} \end{array}$$

In the conference management example, the REVIEWER role requires the capability to read the papers it has been assigned for review and of writing the associated review forms:

$$\begin{array}{lll} \texttt{reads} & \textit{Papers} & \textit{// all the papers it receives} \\ \texttt{changes} & \textit{ReviewForms} & \textit{// one for each of the papers} \end{array}$$

Due to the strict relations between the permission and the environmental models, a graphical representation relating the preliminary roles and their connections with the environment is often a useful outcome of the analysis phase. To this end, Figure 8 details such a graphical representation for the manufacturing pipeline example. Specifically, such an environment-role diagram may help identify inconsistencies between what operations the environment permits and what the agents need (or must be allowed) to do. The relations between the environment and the roles' permissions (identified in Figure 6 by the arrow going from the environmental model box to the preliminary role model box) helps identify several key points about the developing MAS:

- Suppose an environmental resource is available for reading only, whereas one of the identified roles needs to change it. In this case, either the environmental model has to be rethought (if possible) or the capabilities of the roles have to be reduced or re-distributed among roles.

- In physically distributed environments, the diagram helps to determine whether it is reasonable and feasible for a single role to access a wide variety of widely distributed resources, or whether it is more appropriate to divide these capabilities among a set of roles.

- When a role needs access to a variable to which it does not have direct access, a possible solution is to access it via the mediation of a role that does. This may help in identifying the need for a given role to be involved in a specific protocol to gain mediated access to that variable. This inter-dependency of the environmental model and of the preliminary interaction model is denoted in Figure 6 via a proper arrow.

Responsibilities: These attributes determine the expected behavior of a role and, as such, are perhaps the key attribute associated with a role. Responsibilities are divided into two types: *liveness properties* and *safety properties* [36]. Liveness properties intuitively state that "something good happens", i.e., describe those states of affairs that an agent must bring about, given certain conditions. In the conference management example, a liveness property for the REVIEWER role may specify that
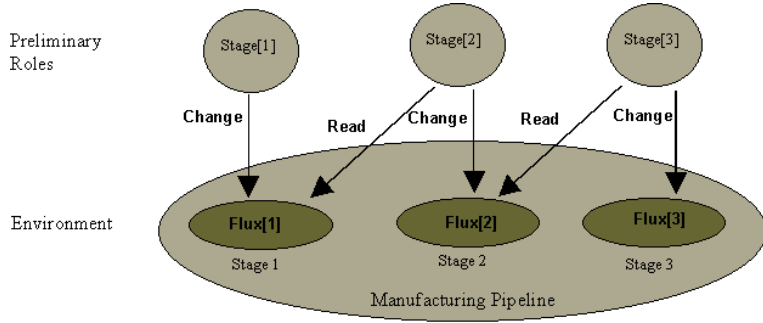
Figure 8: The Environment-Preliminary Roles Diagram

Operator	Interpretation
$x.y$	x followed by y
$x \mid y$	x or y occurs
$x*$	x occurs 0 or more times
$x+$	x occurs 1 or more times
x^{ω}	x occurs indefinitely often
$[x]$	x is optional
$x \parallel y$	x and y interleaved

Table 1: Operators for liveness expressions

a reviewer, on receipt of a paper, has to start the review and eventually produce a completed review. In the manufacturing pipeline example, a liveness property may specify that whenever a substantial change occurs in the flux of items, the STAGE controllers must engage in some coordinated activity to restore a stable condition. In contrast, safety properties are *invariants*. Intuitively, a safety property states that "nothing bad happens", i.e., that an acceptable state of affairs is maintained. In the manufacturing pipeline, the role of STAGE controller must guarantee that the flux of items is always kept within a specific range.

The most widely used formalism for specifying liveness and safety properties is temporal logic, and the use of such a formalism has been strongly advocated for use in agent systems [60]. Although it has undoubted strengths as a mathematical tool for expressing liveness and safety properties, there is always some doubt about the viability of such formal tools for use in everyday software engineering practices. Therefore, without excluding the possibility of using such a formalism, this paper proposes an alternative approach based on regular expressions, as these are likely to be understood by a larger audience. To this end, liveness properties are specified via a *liveness expression* which defines the "life-cycle" of the role (see Table 1). Liveness expressions are similar to the *life-cycle* expressions of FUSION [16], which are in turn essentially regular expressions. Our liveness expressions have an additional operator, "ω", for *indefinite repetitions*.

Liveness expressions, defining the potential execution trajectories through the various activities

and interactions associated with the role, have the general following form:

$$\textsc{RoleName} = \textit{expression}$$

where ROLENAME is the name of the role whose liveness properties are being defined, and *expression* defines the liveness properties of ROLENAME. The atomic components of a liveness expression are either *activities* or *protocols*. An activity is somewhat like a method in object-oriented terms, and corresponds to a unit of action that the agent performs and that does not involve interaction with any other agent. Protocols, on the other hand, are activities that *do* require interaction with others. To give the reader some visual clues, we write protocol names in a sans serif font, and use the same font, underlined, for activity names.

To illustrate liveness expressions, let us consider the responsibilities of a simple REVIEWER role in the conference management example:

$$\textsc{Reviewer} \quad = \quad (\textsf{ReceivePaper}.\underline{\textsf{ReviewPaper}}.\textsf{SendReviewForm})^{\textit{maximum-number}}$$

This expression says that REVIEWER consists of executing the protocol ReceivePaper (for the moment, we treat the protocols simply as labels – we will give a more accurate definition shortly), followed by the activity ReviewPaper and the protocol SendReviewForm. The sequential execution of these protocols and activities is then repeated for the maximum number of papers the role has to deal with. In any case, we emphasize that during the analysis phase it may be impossible to completely determine the liveness expression for a role. For instance, without being committed to a specific organizational structure, one cannot determine if the reception of a paper by a reviewer has to be preceded by some sort of negotiation or bidding. In such cases, the preliminary roles definition can only sketch the activity (i.e., the liveness expression) of a role, to be completed during design.

As an additional example, consider the liveness expression for a role STAGE[i] in the manufacturing pipeline example:

$$\textsc{Stage}[i] \quad = \quad (\underline{\textsf{MonitorFlux}}[i-1]^{\omega}.\textsf{ReduceSpeed}.\textsf{OKReduceSpeed})^{\omega} \parallel \underline{\textsf{ProcessItems}}^{\omega}$$

This expression says that STAGE[i] consists of continuously executing an indefinite number of times the activity MonitorFlux[i-1], to continuously check the flux of incoming items, and to be possibly sometimes interrupted and followed by the execution of the protocols ReduceSpeed and

OKReduceSpeed. These two protocols, intended to negotiate a new speed and actualize the speed change in the pipeline, respectively, have to be executed whenever the safety rule expressing the fact that the input and output fluxes must be the same cannot be respected. In parallel with the above continuous activities, is the – again continuous – ProcessItems activity, devoted to controlling item processing in the stage.

We now turn to safety requirements. These can be specified by means of a list of predicates, typically expressed over the variables listed in a role's permissions attribute. Returning to our REVIEWER role, an agent carrying out this role will be required to ensure that for any received paper it has also received the review form. This can be expressed by the following safety expression:

- *number_of_papers = number_of_review_forms*

By convention, we simply list safety expressions as a bulleted list, each item in the list expressing an individual safety responsibility. It is assumed that these responsibilities apply across *all* states of the system execution. Thus, if the role is of infinitely long duration, the invariants should always be true. However, because something unexpected can always occur in complex systems we prefer to assume a weaker notion for safety expressions; considering them as something that the agent playing a role has to do its best to preserve. In other words, safety expressions indirectly express abnormal situations to which an agent should be able to react. Of course, the need to respect safety expressions may require an agent to tune its liveness properties accordingly. This requirement appears clearly in the manufacturing pipeline example. Here a basic safety rule for the role STAGE[i] it to ensure that the flux of items exceeds for no more than a given threshold T_i (and possibly for no more than a short time interval) that it is able to process:

- *$flux[i] = flux[i-1] \pm T_i$*

To ensure such a property, STAGE[i] must monitor the flux and eventually request a flux reduction, as already identified when discussing the liveness expressions for this role.

With all these definitions in place, it is now possible to precisely define the Gaia roles model. This is used in the analysis phase to define preliminary roles and in the design phase to give the complete specification of all the roles involved in a system. A roles model is comprised of a set of *role schema*, one for each role in the system. A role schema draws together the various attributes discussed above into a single place (Figure 9). In the case of the preliminary role model, such a schema may simply leave some of its parts undefined, to be completed in the architectural design phase. An exemplar instantiation is given for the REVIEWER role in Figure 10.

4.1.4 The Preliminary Interaction Model

This model captures the dependencies and relationships between the various roles in the MAS organization, in terms of one protocol definition for each type of inter-role interaction. Since the roles

Role Schema:	*name of role*
Description	*short description of the role*
Protocols and Activities	*protocols and activities in which the role plays a part*
Permissions	*"rights" associated with the role*
Responsibilities	
Liveness	*liveness responsibilities*
Safety	*safety responsibilities*

Figure 9: Template for role schemata

Role Schema: REVIEWER

Description:
This preliminary role involves receiving papers for review from some conference official, reviewing that paper, and sending back a completed review form.

Protocols and Activities:
ReceivePaper, ReviewPaper, SendReviewForm

Permissions:

	reads	*Papers*	*// all the papers it receives*
	changes	*ReviewForms*	*// one for each of the papers*

Responsibilities

Liveness:

$$\text{REVIEWER} \quad = \quad (\text{ReceivePaper.ReviewPaper.SendReview})^{maximum\_number}$$

Safety:
- *number_of_papers = number_of_reviewforms*

Figure 10: Schema for role REVIEWER

---

model is still preliminary at this stage, the corresponding protocols model must also necessarily be preliminary, for the same reasons.

In Gaia, a protocol can be viewed as an institutionalized pattern of interaction. That is, a pattern of interaction that has been formally defined and abstracted away from any particular sequence of execution steps, to focus attention on the essential nature and purpose of the interaction, rather than on the precise ordering of particular message exchanges (cf. the interaction diagrams of OBJEC-TORY [16, pp198–203] or the scenarios of FUSION [16]).

A protocol definition consists of the following attributes:

- *protocol name*: brief textual description capturing the nature of the interaction (e.g., "information request", "schedule activity X" and "assign task Y");

- *initiator*: the role(s) responsible for starting the interaction;

- *partner*: the responder role(s) with which the initiator interacts;

Figure 11: The ReduceSpeed preliminary protocol definition

- *inputs*: information used by the role initiator while enacting the protocol;

- *outputs*: information supplied by the protocol responder during interaction;

- *description*: textual description explaining the purpose of the protocol and the processing activities implied in its execution.

In the manufacturing pipeline, consider the ReduceSpeed protocol, which forms part of all STAGE[i] roles (Figure 11). This protocol is initiated by one of the stages in the pipeline when it notices that the input flux of items can no longer be tolerated and that a global reduction in pipeline flux is required. At this stage of the development process, however, it is not possible to know which other roles this protocol will involve. In any case, it is possible to identify that the outcome of the protocol will be a new value, properly negotiated, for the flux of items. In the conference management case study, consider the ReceivePaper protocol which is part of the REVIEWER role (Figure 12). This states that the protocol ReceivePaper will be initiated by some role (still undefined at this stage) in charge of assigning papers for review. It involves the role REVIEWER being asked if it can review the proposed paper, eventually leading to REVIEWER receiving the *Paper* (i.e., the specific paper for which the protocol was initiated).

Of course, a more detailed protocol definition (including, for example, the sequencing of messages within the protocol) can be added at any time to complement our rather conceptual definition and provide more specific guidelines for design and implementation. For instance, the AUML notation could serve this purpose well [42].

4.1.5 The Organizational Rules

The preliminary roles and interaction models capture the basic characteristics, functionalities, and interaction patterns that the MAS system must realise, independently of any pre-defined organizational structure. However, as previously stated, there may be general relationships between roles, between protocols, and between roles and protocols that are best captured by organizational rules.

33

Protocol Name: **Receive Paper**		
Initiator: **??** (PC Chair or PC Member)	Partner: **Reviewer**	Input: **Paper info**
Description: When a paper has to be assigned to a reviewer it (by someone undefined at this stage) it will be proposed by sending paper info to one of the potential reviewer		Output: **No, don't review** **OR** **Yes, I review it, send me the full paper**

Figure 12: The ReceivePaper preliminary protocol definition

In Gaia, the perspective on organizational rules is consistent with that on roles' responsibilities: organizational rules are considered as *responsibilities* of the organization as a whole. Accordingly, it is possible to distinguish between safety and liveness organizational rules. The former refer to the invariants that must be respected by the organization for it to work coherently, and the latter express the dynamics of the organization (i.e., how the execution must evolve). In particular:

- *liveness* rules define how the dynamics of the organization should evolve over time. These can include, for example, the fact that a role can played by an entity only after it has played a given previous role, or that a given protocol may execute only after some other protocol. In addition, liveness organizational rules can relate to other liveness expressions belonging to different roles, i.e., relating the way different roles can play specific activities.

- *safety* rules define time-independent global invariants for the organization that must be respected. These can include, for example, the fact that a given role must be played by only one entity during the organization's lifetime or that two roles can never be played by the same entity. In addition, they can relate to other safety rules of different roles or to expressions of the environmental variables in different roles.

Due to their similar nature, organizational rules can be expressed by making use of the same formalism adopted for specifying liveness and safety rules for roles[2]. To illustrate this, consider the following organizational rules, expressing exemplar *liveness* rules:

- $R \rightarrow Q$: means that the role Q can be played by an entity only if it has somewhen earlier played the role R.

- $R^3 \rightarrow Q$: means that the role Q can be played by an entity after it has somewhen earlier played at least three times the role R.

---

[2]The interested reader can refer to [62] for a treatment of organizational rules that makes use of temporal logic and leads to a more formal approach. Again this is not incorporated here for reasons of ease of use by practitioners.

Conversely, safety expressions for organizational rules detail properties that must always be true during the whole life of the MAS:

- $\neg(R|Q)$: means that the two roles R and Q can never be played concurrently by the same entity (recall that the "|" operator for liveness rules expresses concurrency of activities).

- $\neg((R|0)|(Q|0))$: means that the two roles R and Q can never be played alone, i.e., that an agent must play the two roles concurrently;

- $R^{1..N}$: means that the role R must always be played at least once and no more than N times;

- $R(property), R = Role1, Role2, RoleN$: means that the specified *property* has to be spread over all the roles listed.

In the manufacturing pipeline, the correct management of the pipeline requires each of the stage roles to be played only once. This can be expressed by the safety rule:

$$\text{R}^1, \text{R} = (\text{STAGE}[1], \text{STAGE}[2], \ldots, \text{STAGE}[N])$$

As a liveness organizational rule for this example, it is possible to consider the fact that, in addition to the tolerated bias between the flux of items in a single stage, the whole pipeline could be required to have a maximum tolerated bias among all the stages. Such a property, to be intended as a global responsibility of the whole MAS, can be expressed by the following organizational rule:

$$flux[1] = flux[2], \ldots flux[N], \pm T_{global}$$

Turning to the conference management example. The rule expressing the fact that an author cannot act as a reviewer for its own paper can be expressed as:

$$\neg(\text{REVIEWER}(paper(x)) \mid \text{AUTHOR}(paper(x)))$$

The rule that the role of reviewer must be played at least three times for each of the submitted papers (i.e., each must receive at least three reviews) is expressed as:

$$\text{REVIEWER}(paper(i))^{3+}, i = 1, \ldots, number\_of\_submitted\_papers$$

(Such a rule should be coupled with another one avoiding a referee to review a paper multiple times.) The rule expressing the fact that the role in charge of setting up the conference program can only be played after that in charge of collecting all the reviews has been played is expressed as:

$$\text{REVIEWCOLLECTOR} \rightarrow \text{DOPROGRAM}$$

Similarly, liveness and safety organizational rules can be imposed on protocols. For example:

- P^1: the protocol must be executed only once;

- $P(R_1)^1$: the protocol must be executed only once by role R_1;

- $P \rightarrow Q$: the protocol P must necessarily precede the execution of protocol Q;

- $P(R_1) \rightarrow Q(R_1)$: the protocol P must necessarily be executed by role R_1 before R_1 can execute protocol Q;

In the manufacturing pipeline, one of the stages may need to request a reduction in the speed of items because it can no longer operate at the current speed. However, this reduction of the speed (requested by the protocol ReduceSpeed) can only be actualized (e.g., via a protocol OKReduceSpeed) after the protocol ReduceSpeed has involved all the other stages. This can be expressed as:

$$\text{ReqReduceSpeed}^{N-1} \rightarrow \text{OKReduceSpeed}$$

In the conference management example, the rule expressing the fact that the reviewer of *paper*(x) must send one and only one review for that paper, by executing the protocol SendReview, is:

$$\text{SendReview(Reviewer(paper(x)))}^1$$

The rule expressing the fact that only after a reviewer has received a paper for review can it actually send back the review form (by initiating the appropriate protocol) is:

$$\text{ReceivePaper(Reviewer(paper(x))))} \rightarrow \text{SendReview(Reviewer(paper(x)))}$$

The above two rules, together with those previously introduced for expressing the fact that each paper must receive three reviews, ensure that each paper will receive three reviews from different referees. In summary, organizational rules, and their correct identification, are fundamental to the

36

design phase because they constrain the definition of the actual organizational structure population and restrict the number of proper implementations of an organization. In addition, the definition of organizational rules is necessary when the system is open. In this case, organizational rules are important to analyze *(i)* whether and when a new agent can be admitted into the organization and which roles it is allowed to play and *(ii)* which kinds of protocols can be initiated by the members of the organization and when, and which kinds of protocols are an expression of legitimate self-interest. As a final note, the above examples of organizational rules clearly show that collecting and defining the system specifications may also imply discovering which agents have to be present in the organization. Some of the organizational rules for roles are likely to be reflected in the design phase, driving, for example, the number of agents to be instantiated and the number of roles to be played by a single agent. However, although some of these design decisions may become clear well before the design phase, the analysis phase should still disregard the presence of agents and only focus on the more abstract concept of roles.

4.2   The Design Phase: Architectural Design

The output of the Gaia analysis phase systematically documents all the functional (and to some extent non-functional) characteristics that the MAS has to express, together with the characteristics of the operational environment in which the MAS will be situated. These structured specifications have to be used in architectural design to identify an efficient and reliable way to structure the MAS organization, and to complete accordingly the preliminary roles and interactions models.

It is worth emphasizing that, while the analysis phase is mainly aimed at understanding what the MAS will have to be, the design phase is where *decisions* have to be taken about the actual characteristics of the MAS. Therefore, even if some of the activities in design (i.e., the completion of the roles and interactions models) appear to be aimed at refining the outputs of the analysis, they in fact rely on actual decisions about the organizational structure and lead to a modeling of the MAS actual characteristics starting from the specifications.

Of course, any real world project faces the difficult task of identifying when the analysis phase can be considered complete (i.e., mature enough for design decisions to be taken). In most cases, only the initiation of design activities enables missing or incomplete specifications to be identified, or conflicting requirements to be noted. In either case, such findings typically require a regression back to earlier stages of the development process. Gaia is not exempt from such problems, although by making explicit some decisions which are implicit in other methodologies (i.e., the architectural design) it promotes an earlier identification of them.
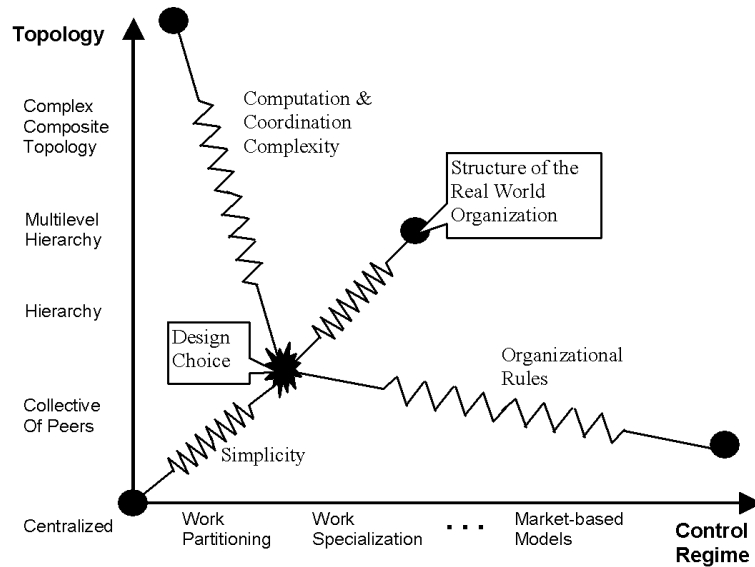
Figure 13: The different forces involved in the identification of the organizational structure

4.2.1 Choosing the Organizational Structure

The choice of the organizational structure is a very critical phase in MAS development, affecting all subsequent phases. Unfortunately, as is always the case with architectural design, it is not possible to identify a precise and formal methodology with which to obtain the "best" design. Nevertheless, a design methodology could and should give guidelines to help the designer make choices.

As Figure 13 illustrates, there are a number of (nearly orthogonal) forces that may drive the identification of an appropriate organizational structure. These include the choice of an appropriate *topology* and of an appropriate *control regime*, among a range of possible choices for the two dimensions. The forces affecting this choice may include: the need to achieve organizational efficiency (or, equivalently, the need for the MAS to properly handle the computation and coordination complexity of a problem); the need to respect organizational rules; and the need to minimize the distance from the real-world organization. All of which may be in tension with the desire to keep the design simple.

Organizational Efficiency and Simplicity: Organizational theory, which has been successfully applied in the analysis and design of distributed software systems [23], states that any member of an organization, whether a human being, a hardware component, or a software agent, exhibits *bounded rationality* [51]. That is, the amount of information it is able to store and process in a given amount of time is limited. Therefore, for an organization to work efficiently it must not overload its members. Whenever the problem of bounded rationality is likely to affect the organization, the organization has to enlarge its resources (i.e., of its members), so as to better distribute the workload

38

among them. The fact that organizations of different sizes may require different organizational topologies derives from the coordination costs that are incurred as the number of members in the organization increases.

As a general guideline (see Figure 13), the simplest organization *topology* that makes it possible to properly handle the computational and coordination complexity should be chosen. At one extreme, a single-member organization experiences no coordination costs and has the advantage of extreme simplicity, although in most cases it is impossible to charge a single agent with all the duties. When an organization is composed of only a few members, the coordination costs can be sufficiently low that collective decisions among all the members are possible. The resulting organizational network is a *collective of peers* in which all members, although possibly committed to different roles, have the same authority over the global organization. When the size of the organization increases still further, the members of the collective can no longer bear the increased coordination costs and a hierarchical topology must be adopted. In the hierarchy, the organization's members are freed from the duty of handling the coordination activities because a *leader* member assumes responsibility for them. As the organization's size increases further, a collective leadership may be needed to handle the increased coordination costs, or a multi-level hierarchy may emerge. In general, any composition of the basic structures can be adopted for the definition of the overall organizational structure. Also, in some cases, hybrid organizational networks can emerge. For example, a collective of peers can be ruled by a leader for a limited portion of its global coordination needs, while maintaining the capability of direct peer-to-peer coordination for others (or, equivalently, a leader in a hierarchy can take charge of a limited set of coordination activities, leaving subordinate members to directly coordinate with one-another for the remaining ones).

The *control regime* handling the interactions between the members of the organization is somewhat orthogonal to the organizational topology. However, in contrast to our simplified representation in Figure 13, it may also somehow influence the organizational efficiency (although in a less significant and direct way than the topology does). In a collective of peers, for example, the control regime can be based either on a *workload partitioning* regime (in which each of the peers has the same role and provides the same services) or on a *workload specialization* (in which each of the peers provides a specific service/activity). The two solutions, depending on the specific application characteristics, may lead to different coordination costs, thus influencing organizational efficiency: when large numbers of similar tasks have to be performed, work partitioning is better at reducing complexity and coordination costs; when the overall workload derives from a few complex activities, work specialization may incur lower coordination costs. Similar considerations may apply to different control regimes applied to different topologies. For instance, it is possible to opt for a hierarchy in which the main goal of the leader is to coordinate the activities of the subordinates so as to obtain a sensible partitioning (i.e., load balancing) of work among them. Alternatively, it is possible to opt for a hierarchy in which each subordinate is specialized, to supply either a specific service or

product, and in which the leader is in charge of the allocation of resources and services so that each subordinate can work at its full load. In the case of very large organizations and when competitive and self-interested behavior are likely to emerge, both the cooperative control regimes of collectives and the authoritative control regimes of hierarchies are often ineffective. This is because of the unavoidable inefficiencies of multi-level hierarchies (or of hierarchies with high fan-out). In such cases, a control regime based on market models for the distribution of workloads and the supply of services needs to be adopted.

In the manufacturing pipeline example, we have already identified (Subsection 3.3) that both a collection of peers (each in charge of a specific piece of work and directly coordinating with each other) and a hierarchy (in which the leader handles all the coordination needs of the stages) are possible. The choice between these alternatives must consider whether the coordination costs between stages can be directly sustained by the stage roles or not. In the conference management example, we have already discussed how the number of papers to be dealt with in the review process – to be sustained by humans of bounded rationality – influences the structure of the real-world organization and, consequently, of the supporting agent-based systems.

As an additional note, it is worth pointing out that additional forces could influence the choice of the organizational structure and, specifically, its topology. For instance, a problem which is intrinsically distributed in physical space (as in the manufacturing pipeline) and in which the resources of the environment cannot be accessed by every agent from everywhere at the same costs, may mandate the adoption of certain specific distributed topologies.

The Influence of Organizational Rules: Another factor impacting on the choice of the organizational structure (primarily of the control regime, as from Figure 13, although not exclusively of it) is the need for the MAS to respect organizational rules and to be able to enact them during execution.

Some types of organizational rules, typically safety ones, may express very direct and intuitive constraints directly driving the adoption of specific organizational structures (e.g., constraints specifying that two roles have to be played by the same agent or that the same agent cannot play two specific roles together). Clearly, in these cases, and whatever the specific topology adopted, the control regime must be tuned so as to distribute tasks in the organization in respect of the constraints. In the conference management example, the fact that a submitted paper has to be reviewed by at least three different reviewers simply rules out the possibility, during review assignments, to consider a work partitioning regime for the sake of load balancing.

The most significant case in which organizational rules impact on the organizational structure is in the presence of competitive and self-interested behaviour. In such cases, not only is it necessary that the adopted structure of a system respects the organizational rules, but also that the adopted structure promotes their active enactment. In fact, on the one hand, agents that dynamically enter an organization may be unaware of the local rules, and, on the other hand, self-interested agents

may strive to opportunistically ignore such rules to achieve their own specific aims. In such cases, the organizational structure and specifically the control regime of interactions must be shaped so as to make it possible to enact some form of control over the execution of foreign and self-interested agents (or, equivalently, to control which agents can play a specific role and how they play it) so that the organizational rules are respected overall. In the case of a big conference, for example, the need for the PC Members to evenly and efficiently distribute papers to reviewers suggests the adoption of a market-inspired bidding mechanism to rule the activities at the lowest level of the multilevel hierarchy. This solution, potentially the most efficient one to implement review assignments, would rely on making the list of papers public and on having papers assigned to those reviewers expressing their availability and willingness to review them. However, to avoid a reviewer handling its own papers (and, more generally, not to leave a reviewer a free choice about which papers to review) the control regime cannot be based solely on this mechanism. Rather, the bidding process has to be considered only as a preliminary phase, to guide the final and authoritative decisions of the PC members about review assignments.

The above discussion also highlights the fact that the enforcement of some organizational rules may have a significant effect on the computational or coordination costs (i.e., in the above example, the need to control review assignment charges PC Members of the additional tasks of supervising the bidding process). Thus, organizational rules may impact on the choice of the organization topology. For instance, consider a liveness rule that forces a specific ordering on the execution of a set of interaction protocols involving a group of agents. In this case, the adoption of a simple collective of peers as a basic topology would impose so much load on each of the agents to control the enactment of the rule that a designer would do better to opt for a hierarchy, with a leader in charge of controlling the protocols' execution sequence. A situation of this kind may also occur in the manufacturing pipeline. Here there is a need to ensure that all stages have executed the `OKReduceSpeed` protocol before the `ReduceSpeed` one can be executed. This may, in turn, be the primary force suggesting the adoption of a hierarchical topology instead of a collective one.

The Influence of the Real-World Organization: In many cases, a MAS is intended to support the structure of a real-world organization. Therefore, in general, the structure of the real-world organization acts as an attracting force in the definition of the organizational structure (as shown in Figure 13), with a natural tendency to define it so as to mimic/reflect the real-world structure. Notable examples are agent-based systems for computer supported cooperative work, in which agents are associated with each of the team members and in which agents usually have to cooperate according to the interaction patterns of the real-world organization. In a similar way, in agent-based systems for electronic commerce, the patterns of interactions between the agents and their organizational roles are likely to mimic the ones that can be found in human commercial transactions.

Having the structure of the MAS reflect the structure of the real-world organization may be

important for the sake of conceptual simplicity, and in several cases this may also come as a requirement, at least for some portions of the structure. However, is not always the case that this attracting force should be accommodated.

Since the construction of a MAS necessarily implies an in-depth analysis of the corresponding real-world organization, it often happens that this process uncovers inefficiencies in the real-world organization. These inefficiencies may simply be related to the fact that the organizational structure is not appropriate for the size and objectives of the organization, or it may be more subtle. The latter case includes the discovery of role underloads (e.g. the duties of a member of the organization are significantly lower than those it could cope with) or role ambiguities (e.g., some members of the organization do not have a well-defined role or they misunderstand the responsibilities). In such cases, and when the designer of the system does not have the authority to modify and improve the real-world organization, it is still possible to improve its software counterpart.

In addition, the availability of software systems to support the work of teams and organizations tends to strongly impact on the way of working and on the real-world organizational structure itself, even if it were previously efficient (a phenomenon that applies to all information technology placed in this situation). Once the members of the organization can have part of the work delegated to agents, their role in the organization can become underloaded, and the whole real-world organization may be forced to re-think its structure to face the reduced computational and coordination costs in a more efficient way. Where possible, this problem has to be taken into account in the design of the software system. However, since the problem is most likely to show its effect over time, after the MAS has entered its life in the real-world organization, the designer may be forced to tune the chosen organizational structure of the MAS to these new needs.

The conference management example exemplifies the above influences well. For instance, the PC Chair in charge of "manually" assigning papers to reviewer PC members (i.e., for each and every paper, by finding at least three PC members to act as reviewer for that paper, as per Figure 4) may be overwhelmed by this work as the number of papers becomes large. Therefore, as is usual in large conferences, the PC Chair is forced to simply partition the papers among PC members (as per Figure 5) and have the PC members help themselves to find the needed number of reviewers for their assigned papers (i.e., delegating to PC members the role REVIEWCATCHER). However, when an agent-based system is provided to support the conference organization committee, it may be the case that the work of the PC Chair and of the PC members dramatically reduces and, for instance, the PC Chair may be able to directly recruit reviewer PC members for a much larger number of papers.

4.2.2 Exploiting Organizational Patterns

Whatever the specific factors that may determine the choice of an organizational structure, it is likely that similar issues have been faced in the past, and have led to similar considerations and

choices. This is the stage where the availability of a catalogue of possibly modular and composable "organizational structures", describing in much more detail than we have done how and when to exploit a specific structure, will greatly help the designer [49, 31, 14] and will effectively complement our general guidelines. In fact, the guidelines we have described are intended to be of help in evaluating the right trade-off in a rather coarse-grained taxonomy of structures and control regimes. However, they consider neither a number of variations on the theme that could be conceived, nor a number of non-functional aspects possibly affecting such a more detailed choice (e.g., mobility, security, resilience, etc.). The specific identification of these variations and their impact on specific non-functional aspects may be well supported by a structured catalogue of organizational structures. This will help the designer to re-use both well-documented and motivated design choices and the design work related to the representation of such a structure.

4.2.3 Representing the Organizational Structure

Once the identification of a suitable organizational structure for the MAS is complete, the designer has to determine how to effectively represent it. In this context, we suggest the coupled adoption of a formal notation and of a more intuitive graphical representation.

The obvious means by which to formally specify an organization is to explicate the inter-role relationships that exist within it (to represent the topology of the organizational structure) and their types (to represent the control regime of the organization). We emphasize that there is no universally accepted ontology of organizational relationships. Nevertheless, as a first pass towards a more complete characterization and formalization, we can identify certain types of relationships that are likely to frequently occur in MAS organizations. To give a few examples: a *control* relationship may identify an authority relationship of one role over another, in which a role can (partially) control the actions of another role; a *peer* relationships may express the fact that two roles have equal status; a *dependency* relation may express the fact that one role relies on some resources or knowledge from other roles for its accomplishment. In any case, these exemplar relationship types are neither mutually exclusive (e.g., a control relationship may also imply a dependency relationship and a dependency relationship may exist between peers), nor claim to define a complete set (other types of relationships can be identified).

In the manufacturing pipeline, and assuming that a collective organization has been chosen, all the stages are peers. This can be simply expressed as:

$$\forall i,j \text{ s.t. } i \neq j, \text{STAGE}[i] \xrightarrow{peer} \text{STAGE}[j]$$

If the hierarchical organization is chosen instead, the coordinator is in charge of supporting all the stages for the protocol of speed negotiation. This structure expressing the dependencies of stages on the coordinator to negotiate the pipeline flux can be represent as:

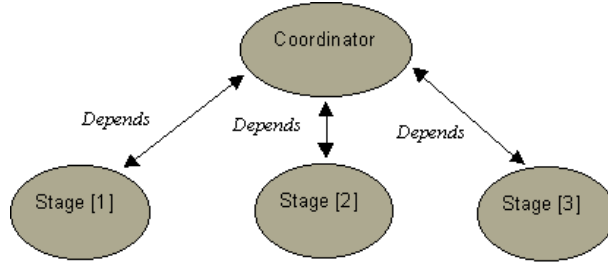$$\forall i, \text{STAGE}[i] \xrightarrow{depends\_on} \text{COORDINATOR}$$

Figure 14: The manufacturing pipeline: representation of the hierarchical organizational structure

In the conference management example, and assuming the choice of a multilevel hierarchy, the representation of the organizational structure is given below. The relations, respectively, express the fact that the PC Chair controls the papers assigned to the REVIEWCATCHER roles, that the REVIEWCATCHER role controls REVIEWERS, and that REVIEWCATCHERS are peers with respect to the negotiation of papers.

$$\forall i, \text{PC CHAIR} \xrightarrow{control} \text{REVIEWCATCHER}[i]$$

$$\forall i, j, \text{REVIEWCATCHER}[i] \xrightarrow{control} \text{REVIEWER}[j]$$

$$\forall i, j, \text{REVIEWCATCHER}[i] \xrightarrow{peer} \text{REVIEWCATCHER}[j]$$

For all the above representations to be useful to developers, they must be properly detailed with a description of the semantics of the relations (i.e., of the protocol and activities involved in such relations, as detailed in the following subsection) and with any additional textual comments that could possibly enrich the notation and make it more comprehensible.

Since the formal representation of the relationship may not be sufficiently intuitive for all practitioners, a graphical representation may be used to complement it. Such a representation can simply reduce to representing roles in terms of "blocks", connected by annotated arrows, to represent relations and their types. Such a representation for the manufacturing example and the hierarchical organization is given in Figure 14. Coupling such a graphical representation with the one of the environmental model (representing in an abstract way the operational environment in which the MAS will be situated) can lead to a complete and intuitive graphical representation of all the entities involved in the MAS, of their interactions with each other, and with the operational environment.

Again, we emphasize that Gaia, per se, is not intended to commit to the adoption of a specific notation. With regard to organizational structure, besides the simple representational methods described above, other notations and graphical representation can be adopted to describe and represent roles and their interactions (e.g., AUML diagrams [4]).

4.2.4   Completion of Role and Interaction Models

Once the organizational structure is defined, the preliminary role and interaction models (as identified in the analysis phase) can be transformed into complete roles and interactions models (describing in detail the characteristics of all roles and stages involved in the MAS that will actually be produced). In fact, with the organizational structure identified, the designer knows which roles will have to interact with which others (as derived from the organization topology) and which protocols will have to be executed (as derived from the control regime of the organization). Thus, the completion of the role and protocol model amounts to:

- defining all the activities in which a role will be involved, as well as its liveness and safety responsibilities;

- defining *organizational roles* – those whose presence was not identified in the analysis phase, and whose identification derives directly from the adoption of a given organizational structure;

- completing the definition of the protocols required by the application, by specifying which roles the protocol will involve;

- defining *organizational protocols* – those whose identification derives from the adopted organizational structure.

The complete definition of the roles and protocols can exploit the representation already introduced with respect to the preliminary roles and protocols (see Figures 9, 10, 11, 12).

With regard to the already identified preliminary roles and protocols, the difference is that the complete representation now includes all the characteristics of roles and protocols (i.e., for roles the full identification of activities and services, and for protocols the complete identification of the involved roles). As a simple example, Figure 15 shows how the preliminary ReduceSpeed protocol (represented in Figure 11) can be completed once a hierarchical organizational structure is chosen (so that the partner of any STAGE[I] initiating the protocol can be identified in the CONTROLLER leading the hierarchy).

With regard to roles and protocols whose definition directly derives from the adoption of a specific structure, their representation has to be defined from scratch (or by exploiting templates from catalogues of organizational patterns). As a simple example, Figure 16 shows the definition of a protocol NegotiateSpeedReduction for the manufacturing pipeline application, whose introduction derives from having chosen a hierarchical organization: the CONTROLLER, once having been contacted by a STAGE[I] asking for a speed reduction, has to negotiate an acceptable new speed with all the other stages.

When completing the preliminary role and interaction models, it is important to clearly preserve the distinction between those characteristics that are "intrinsic" (i.e., independent of the use of
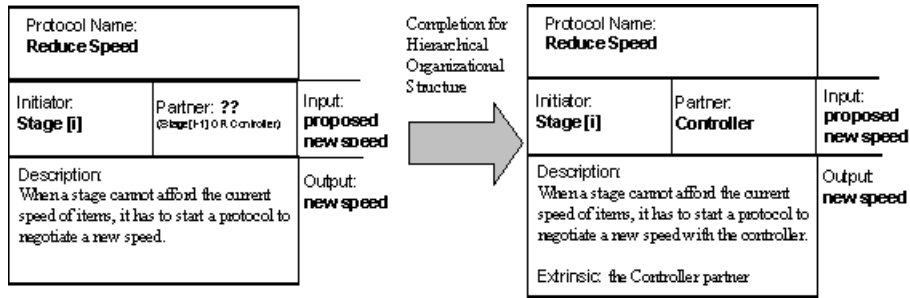
Figure 15: The completion of the ReduceSpeed protocol

Figure 16: The definition of the NegotiateSpeedReduction protocol

the role/protocol in a specific organizational structure) and those characteristics that are "extrinsic" (i.e., derive from the adoption of a specific organizational structure). For instance, in the manufacturing pipeline example, the fact that the STAGE[i] roles initiate the ReduceSpeed protocol is intrinsic, whereas the fact that such a protocol involves the COORDINATOR role is extrinsic (since the COORDINATOR role is an organizational role, whose integration in the system derives from the choice of a hirarchical organizational structure). The above separation, that we have simply preserved as a comment to the description of the protocols (as in Figures 15 and 16), assumes a fundamental importance in the perspectives of re-use and design for change. If the definition of a role clearly specifies the logical separation between the two parts, and the implementation of the role somehow does the same, this enables the role be re-used in different systems, independently of the specific organizational structure. This distinction also means that changes in the organizational structure can be made without forcing system developers to re-design and re-code agents from scratch. Furthermore, if the implementation of the role, other than maintaining a clear separation between intrinsic and extrinsic characteristics also supports dynamic changes in the agent structure (as, for example, in frameworks supporting the run-time modification of the roles played by agents [10]), the approach could promote the definition and implementation of systems capable of dynamically changing their internal organizational structure – by having agents re-adapt their extrinsic part to fit into a new organizational structure – in response to changed conditions. However, dealing with

46

such dynamic re-organizations would need the multiagent system to monitor its execution and take decisions about its re-structuring and these are issues that are beyond the scope of this paper and of the current version of Gaia.

Once completed, the roles and interaction models represent an operational description of the MAS organization that could be effectively exploited, possibly with the support of the environmental model (to get a better idea of the characteristics of the MAS operational environment), for the detailed design of the MAS.

4.3 The Design Phase: Detailed Design

The detailed design phase is responsible for eventually identifying the agent model and the services model which, in turn, act as guidelines for the actual implementation of agents and their activities.

4.3.1 Definition of the Agent Model

In the Gaia context, an agent is an active software entity playing a set of agent roles. Thus, the definition of the agent model amounts to identifying which agent classes are to be defined to play specific roles and how many instances of each class have to be instantiated in the actual system.

Typically, as outlined in Subsection 4.2.1, there may be a one-to-one correspondence between roles and agent classes: a given role will be played by an instance of an agent of a class in charge of implementing that role. Such a correspondence naturally derives from the methodological process that was adopted in which, especially during the identification of the organizational structure, the concept of role has implicitly assumed ever greater concreteness. However, given that (*i*) the organizational efficiency is not affected, (*ii*) bounded rationality problems do not emerge, and (*iii*) this does not violate organizational rules, a designer can choose to package a number of closely related and strongly interacting roles in the same agent class for the purposes of convenience. There is obviously a trade-off between the *coherence* of an agent class (how easily its functionality can be understood), the efficiency considerations that come into play when designing agent classes, and the need to minimize mismatches with the real-world organization that the system intends to support.

With reference to the conference management example, it may be clear that it is desirable to have the same agent playing the role of REVIEWCATCHER and REVIEWCOLLECTOR. Such a choice affects neither the efficiency nor the bounded rationality of the agent, because the two roles are played at different times. In addition, such a choice does not conflict with any organizational rules. In contrast, such a choice may compact the whole design by reducing the number of agent classes and instances and it may also minimize the conceptual complexity (in that also in the real-world organization the two roles are typically played by the same person). With reference to the manufacturing pipeline, the choice of a one-to-one mapping of roles to agent classes is the most natural and it is not easy to think of a reasonable alternative. Thus, the system will be described by a STAGE

class and by a COORDINATOR class, whose instances will be in charge of implementing the roles with the same names.

The agent model of Gaia can be defined using a simple diagram (or table) specifying, for each agent class, which roles will map to it. In addition, the agent model can document the instances of a class that will appear in the MAS. This can be done by annotating agent classes with qualifiers from FUSION [16]. An annotation n means that there will be exactly n agents of this class in the run-time system. An annotation $m..n$ means that there will be no less than m and no more than n instances of this class in a run-time system ($m < n$). An annotation $*$ means that there will be zero or more instances at run-time, and $+$ means that there will be one or more instances at run-time.

With reference to the conference management example, the notation:

$$\text{PC CHAIR}^1 \xrightarrow{play} \text{REVIEWCATCHER, REVIEWCOLLECTOR}$$

simply expresses that the agent class PC CHAIR will be defined to play both role REVIEWCATCHER and role REVIEWCOLLECTOR, and that a single instance of this class will be created in the MAS. In the case of the manufacturing pipeline, the notations:

$$\text{STAGE}^N \xrightarrow{play} \text{STAGE}$$

$$\text{COORDINATOR}^1 \xrightarrow{play} \text{COORDINATOR}$$

express the fact that the will be N agents (one for each of the N stages of the pipeline) of the class STAGE to play the STAGE role, and (if a hierarchical organizational structure has been chosen) a single agent of the class COORDINATOR to play the COORDINATOR role.

It is worth highlighting that the agent model of Gaia is of use in driving the static assignment of roles to agent classes, in which the implementation of the agent class is assumed to hardwire the code to implement one or more roles. Of course, if the implementation adopts a framework enabling agents to dynamically assume roles (as, for example, in the BRAIN framework [10]), the concept of agent class of Gaia (and so the Gaia agent model) would no longer be of use.

As a final note, inheritance is given no part in Gaia's agent models, since our experience is that a MAS typically contains only a comparatively small number of roles and classes, making inheritance of little use in design. Of course, when it comes to actually *implementing* agents, inheritance will be used in the normal object-oriented fashion.

4.3.2 The Services Model

As its name suggests, the aim of the Gaia services model is to identify the *services* associated with each agent class or, equivalently, with each of the roles to be played by the agent classes. Thus, the services model applies both in the case of static assignment of roles to agent classes as well as in the case where agents can dynamically assume roles.

While in OO terms a service would roughly correspond to a method, agent services are not available for other agents in the same way that an object's methods are available for another object to invoke (as discussed in section 2.2). Rather, due to the intrinsically active nature of an agent, a service is better thought of as a single, coherent block of activity in which an agent will be engaged. Such a service does not have to necessarily be triggered by external requests (i.e., the agent is autonomous in its activities), neither does an external request necessarily trigger a service (autonomy also implies there are internal decision capabilities).

For each service that may be performed by an agent, it is necessary to document its properties. Specifically, we must identify the *inputs*, *outputs*, *pre-conditions* and *post-conditions*. Inputs and outputs to services will be derived in an obvious way from both the protocols model (for services involving the elaboration of data and knowledge exchange between agents) and the environmental model (for services involving the evaluation and modification of environmental resources). Pre- and post- conditions represent constraints on the execution and completion, respectively, of services. These are derived from the safety properties of a role, as well as from the organizational rules, and may involve constraints on the availability and the specific values assumed by either environmental resources or data and knowledge from other agents.

The services that compose an agent are derived from the list of protocols, activities, responsibilities and liveness properties of the roles it implements. At one extreme, there will be at least one service for each parallel activity of execution that the agent has to execute. However, even for sequential activities of execution, there may be a need to introduce more services to represent different phases of the agent execution. Let us return to the conference management example and, specifically, to the REVIEWER role. There are three activities and protocols associated with this role: ReceivePaper, ReviewPaper, and SendReview. As these protocols and activities are to be performed sequentially during the agent execution, it is possible to think of defining the agent in terms of a single service, to be activated upon request (i.e., upon inititation by a REVIEWCATCHER of the ReceivePaper protocol), and eventually leading to the execution of the SendReview protocol. The input of the service is a request for reviewing a paper. The output of the service is the completed review form. The pre-conditions on the execution of the service could be that the same agent has not already executed that service *maximum_number* times (i.e., is not already reviewing the maximum allowed number of papers) and that the paper is available for reading in the environment. The post-condition is that the review form has been correctly completed. However, since the outcome of the ReceivePaper protocol may be different (e.g., a reviewer may decline to review a paper), the design would be better to split the larger service into two smaller ones. This may separate the service associated with the execution of the ReceivePaper protocol from the one associated with the paper review, the latter being executed after a paper has been accepted for review.

The Gaia services model does not prescribe an implementation for the services it documents. Rather the developer is free to realise the services in any implementation framework deemed ap-

propriate. For example, it may be decided to implement services directly as methods in an object-oriented language. Alternatively, a service may be decomposed into a number of methods.

4.4 Outcome of the Design Phase

After the successful completion of the Gaia design process, developers are provided with a well-defined set of agent classes to implement and instantiate, according to the defined agent and services model. As already stated, Gaia does not deal with implementation issues and considers the output of the design phase as a specification that can be picked up by using a traditional method or that could be implemented using an appropriate agent-programming framework. These considerations rely on the assumption that specific technological constraints (e.g., those that may occur with regard to modeling and accessing to the environment) are identified earlier in the process.

As an additional note, although the above methodology is explicitly conceived for open agent systems, its outcomes can also be exploited by:

- third-party developers of agents, whenever they intend to have their agents execute in other organizations, as can be the case of agents that are used in agent-based marketplaces;

- agent users, whenever they wish to exploit their personal assistant agents in the context of specific workgroup applications.

In both cases, assuming that the developers of the existing MAS have made all the Gaia models publicly available, an agent developer will be able to check whether their agent can enter the system and, if so, what rules and procedures it needs to conform to.

5 Related Work

Research in the area of agent-oriented software engineering has expanded significantly in the past few years. Several groups have started addressing the problem of modeling agent systems with appropriate abstractions and defining methodologies for MASs development.

5.1 Modeling Abstractions

Several researchers in this area have defined modeling techniques specifically tuned to the abstractions promoted by agents and MASs (as discussed in section 2.1), without attempting to define complete methodologies for the development of MASs.

As stated in Section 1, we believe conventional analysis and design methodologies, such as object-oriented ones [7], are ill suited to MASs because of the fundamental mismatch between the abstractions they provide [59]. Consequently, we believe that those efforts in the area of agent-based

computing that attempt to describe agents and MASs by simply *applying* well-assessed object-oriented modeling techniques (e.g. [29]) will inevitably fall short. For instance, although it may be possible to represent concepts such as organizational rules and complex interaction protocols in terms of the abstractions of, say, object-oriented computing, this is a convoluted process and one that lowers the abstraction level. Nevertheless, we are aware that the adoption of well-known notations and modeling techniques might facilitate the rapid diffusion of agent-based computing as a software engineering paradigm.

For these reasons, we feel that a more promising approach is the attempt to *extend* well-assessed abstractions and modeling techniques. For instance, the AUML effort [42, 4] starts from UML [46] and extends it with additional abstractions and notations specifically tuned to the particular problems of agent-based applications. However, to date, AUML has only achieved good results in modeling complex interaction protocols between components. It still lacks appropriate models for complex organizational structures and the associated organizational laws [4]. Nevertheless, AUML is gaining acceptance in the area and a large amount of work is being devoted to finding suitable ways of extending it further to make it more suitable. Therefore, as it matures, AUML is likely to become a useful companion to Gaia.

Besides these object-oriented notations, a number of agent-specific modeling abstractions and techniques have been proposed in recent years (see [26] for a survey, and [56] for a discussion). Several of these attempt to exploit the idea of a MAS as a computational organization. In most cases, these proposals define an organization simply as a collection of roles (i.e., a role model). This is what happens, for example, in the AALAADIN system [21], which models MASs in terms of organizations where "the group structure", characterizing organizations, is simply the collection of roles that compose them. Analogously, in the ToolKit approach [18], an organization is defined simply by the set of roles that compose it and by the interaction protocols that have to occur between roles. Neither of these approaches incorporate the notions of organizational rules or organizational structures and, for the reasons we have already outlined in Section 3, will be limited in the range of agent systems they can deal with.

A limited amount of work explicitly addresses the problem of defining organizational rules to orchestrate the behavior of a multiagent organization. However, one application area in which there is a particularly strong need for organizational rules is that of computational markets. For systems in which heterogeneous agents meet to perform commercial transactions, it is important to identify and enact rules to guarantee that the overall activity of the market can correctly progress, despite possibly opportunistic behavior by participants [19]. Such considerations were behind the implementation of the Fishmarket framework for agent-mediated auctions, for example [41]. Here, there is a need to compel agents to act in accordance with the social conventions that rule the organization of an auction. This is then implemented in terms of controller agents associated with each of the agents participating in an auction. Other agent-based approaches for Web-based workgroup applications

51

start from similar considerations [15, 12], and conceive interactions between agents as occurring via "active environments" whose behavior can be modeled so as to implement specific policies for governing agent interactions. In other words, active environments are exploited as the repository of organizational rules. Our work differs from these endeavors in that we focus on the particular organizational abstractions that are appropriate for the software engineering process, rather than on the specific technologies to implement them.

5.2 Agent-Oriented Methodologies

Several complete methodologies for the analysis and design of MASs have been proposed so far (see [50] for a survey). However very few of them explicitly focus on organizational abstractions.

The MASE Methodology [55] provides guidelines for developing MASs based on a multi-step process. In analysis, the requirements are used to define use-cases and application goals and sub-goals, and eventually to identify the roles to be played by the agents and their interactions. In design, agent classes and agent interaction protocols are derived from the outcome of the analysis phase, leading to a complete architecture of the system. However, the MASE process fails to identify any organizational abstraction other than the role model: there are no abstractions exploited in the analysis phase that can be assimilated to organizational rules, and the definition of the organizational structure is derived as an implicit outcome of the analysis phase. As explicitly acknowledged by the MASE designers, this makes MASE suitable only for closed agent systems.

The MESSAGE methodology [13] exploits organizational abstractions that can be mapped into the abstractions identified by Gaia. In particular, MESSAGE defines an organization in terms of a *structure*, determining the roles to be played by the agents and their topological relations (i.e., the interactions occurring among them). This corresponds to the concept of organizational structure promoted by Gaia. In addition, in MESSAGE, an organization is also characterized by a *control entity* and by a *workflow structure*. These two concepts, when taken together, map into Gaia's concept of organizational rules and they determine the laws that agents' actions and interactions have to conform to during execution. Still, MESSAGE does not address the problem of identifying and explicitly modeling the organizational rules (they are implicitly woven into the organizational structure) and it does not recognize the need to explicitly design the organizational structure (that is assumed as an implicit outcome of the analysis phase). As an additional note, MESSAGE is tightly bound to UML (and to AUML), while Gaia is more general, and provides clear guidelines without committing to any specific modeling techniques.

The Gaia methodology described in this paper is an extension of the version described in [59]. The first version of Gaia, as is the case with the current one, provided a clean separation between the analysis and design phases. However, as already noted in this paper, it suffered from limitations – similar to the ones of MESSAGE and MASE – caused by the incompleteness of its set of abstractions. The objective of the analysis phase in the first version of Gaia was to define a fully

elaborated role model, derived from the system specification, together with an accurate description of the protocols in which the roles will be involved. This implicitly assumed that the overall organizational structure was known a priori (which is not always the case). In addition, by focusing exclusively on the role model, the analysis phase in the first version of Gaia failed to identify both the concept of global organizational rules (thus making it unsuitable for modeling open systems and for controlling the behavior of self-interested agents) and the modeling of the environment (which is indeed important, as extensively discussed in this paper). The new version of Gaia overcomes these limitations.

A further extension to the first version of Gaia is also proposed in [28], and is aimed at addressing the same limitations (i.e., representation of the environment and organizational rules). However, that proposal is still rather preliminary and lacks a coherent integration in the Gaia process.

The TROPOS methodology, first proposed in [8] and refined in [31], shares with Gaia both the adoption of the organizational metaphor and an emphasis on the explicitly study and identification of the organizational structure. Like Gaia, TROPOS recognizes that the organizational structure is a primary dimension for the development of agent systems and that an appropriate choice of it is needed to meet both functional and non-functional requirements. However, TROPOS also differs from Gaia in several aspects. On the one hand, the TROPOS methodology defines uniform and coherent guidelines for the activities of both early and late requirements engineering, while Gaia's analysis mostly deals with late requirements engineering. However, as already stated, we do not exclude the possibility of integrating in Gaia early requirements analysis, possibly adapting (as in TROPOS) methods and techniques from goal-oriented analysis [40, 14]. On the other hand, TROPOS's analysis does not explicitly identify the concept of organizational rules. Although some explicit structural dependencies between roles are required to be identified in the analysis phase, no TROPOS model is able to capture global laws that apply to multiple organizational roles or to the organization as a whole. As already discussed in the paper, this may undermine the effectiveness of the analysis and increase the complexity of the subsequent design phase.

6 Conclusions and Future Work

This paper has argued that agent-oriented computing is an appropriate software engineering paradigm for the analysis, design, and development of many contemporary software systems. In particular, we focused on the key issues related to the identification of appropriate abstractions for agent-based software engineering and to the definition of a suitable methodology for the analysis and design of complex applications in terms of multiagent systems. In so doing, the main contributions of this paper are:

- a clarification of the relationship between the agent-oriented approach to software development and that of more traditional (e.g., object- and component-based) approaches;

- the identification of a suitable set of core abstractions, inspired by an organizational metaphor, to be used during the analysis and design of multiagent systems; and

- the development of a clear methodology, centered around organizational abstractions, for the analysis and design of open multiagent systems.

While this is an important step towards the widespread acceptance of agent-based computing as a software engineering paradigm much work remains to be done. In particular:

- robust tools, programming models and development environments that embody the appropriate set of agent abstractions need to be made widely available and easy to use by software engineering practitioners;

- catalogues of organizational patterns (to guide the definition of the organizational structure) and libraries of re-usable design components (to guide the detailed design of agents) need to be developed and disseminated in a widely accessible form;

- while the presented Gaia methodology suggests a rather sequential approach to software development, proceeding linearly from analysis to design and implementation, software development have often to deal with unstable requirements, wrong assumptions, and missing information. This requires the designer to step back from the current activities and, perhaps, to re-think some of the decisions. As it stands, however, there are no explicit structures or processes in Gaia for doing this.

References

[1] H. Abelson, D. Allen, D. Coore, C. Hanson, G. Homsy, T. Knight, R. Napal, E. Rauch, G. Sussmann, and R. Weiss. Amorphous computing. *Communications of the ACM*, 43(5):43–50, May 2000.

[2] O. Babaoglu, H. Meling, and A. Montresor. Anthill: A framework for the development of agent-based peer-to-peer systems. In *Proceedings of the 22nd International Conference on Distributed Computing Systems*, pages 15–22. IEEE CS Press, Vienna (A), July 2002.

[3] L. Bass, P. Clements, and R. Kazman. *Software Architectures in Practice (2nd Edition)*. Addison Wesley, Reading (MA), 2003.

[4] B. Bauer, J. P. Muller, and J. Odell. Agent uml: A formalism for specifying multiagent software systems. *International Journal of Software Engineering and Knowledge Engineering*, 11(3):207–230, April 2001.

[5] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.

[6] E. Bonabeau, M. Dorigo, and G: Theraulaz. *Swarm Intelligence. From Natural to Artificial Systems*. Oxford University Press, Oxford (UK), 1999.

[7] G. Booch. *Object-oriented Analysis and Design (second edition)*. Addison Wesley, Reading (MA), 1994.

[8] P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia, and J. Mylopoulos. A knowledge level software engineering methodology for agent oriented programming. In *Proceedings of the 5th International Conference on Autonomous Agents*, pages 648–655. ACM Press, Montreal (CA), June 2001.

[9] S. Bussmann. Agent-oriented programming of manifacturing control tasks. In *Proceedings of the 3rd International Conference on Multi-Agent Systems*, pages 57–63. IEEE CS Press, Paris (F), June 1998.

[10] G. Cabri, L. Ferrari, and L. Leonardi. Enabling mobile agents to dynamically assume roles. In *Proceedings of the 2003 ACM Symposium on Applied Computing*, pages 56–60. ACM Press, Melbourne (FL), March 2003.

[11] G. Cabri, L. Leonardi, and F. Zambonelli. Mobile-agent coordination models for internet applications. *IEEE Computer*, 33(2):52–58, February 2000.

[12] G. Cabri, L. Leonardi, and F. Zambonelli. Engineering mobile agent applications via context-dependent coordination. *IEEE Transactions on Software Engineering*, 28(11):1034–1051, November 2002.

[13] G. Caire, W. Coulier, F. Garijo, J. Gomez, J. Pavon, F. Leal, P. Chainho, P. Kearney, J. Stark, R. Evans, and P. Massonet. Agent-oriented analysis using message/uml. In *Proceedings of the 2nd International Workshop on Agent-Oriented Software Engineering*, volume 2222 of *LNCS*, pages 119–135. Springer Verlag, 2002.

[14] J. Castro, M. Kolp, and J. Mylopoulos. Towards requirements-driven information systems engineering: the tropos project. *Information Systems*, 27(6):365–389, June 2002.

[15] P. Ciancarini, A. Omicini, and F. Zambonelli. Multiagent systems engineering: the coordination viewpoint. In *Intelligents Agents VI: Agent Theories, Architectures, and Languages*, volume 1767 of *LNAI*, pages 250–259. Springer Verlag, 2000.

[16] D. Coleman, P. Arnold, S. Bodoff, D. Dollin, H. Gilchrist, F. Hayes, and P. Jeremas. *Object-Oriented Development: the FUSION Method*. Prentice-Hall International, Hemel Hampstead (UK), 1994.

[17] A. Collinot, A. Drogoul, and P. Benhamou. Agent-oriented design of a soccer robot team. In *Proceedings of the 2nd International Conference on Multi-agent Systems*, pages 41–47. IEEE CS Press, Kyoto (J), 1996.

[18] Y. Demazeau and A. C. Rocha Costa. Populations and organizations in open multi-agent systems. In *Proceedings of the 1st National Symposium on Parallel and Distributed AI*. 1996.

[19] M. Esteva, J. A. Rodriguez-Aguilar, C. Sierra, P. Garcia, and J. L. Arcos. On the formal specifications of agent institutions. In *Agent-Mediated Electronic Commerce*, volume 1991 of *LNCS*, pages 126–147. Springer Verlag, 2001.

[20] D. Estrin, D. Culler, K. Pister, and G. Sukjatme. Connecting the physical world with pervasive networks. *IEEE Pervasive Computing*, 1(1):59–69, 2002.

[21] J. Ferber and O. Gutknecht. A meta-model for the analysis and design of organizations in multi-agent systems. In *Proceedings of the 3rd International Conference on Multi-Agent Systems*, pages 128–135. IEEE CS Press, Paris (F), June 1998.

[22] I. Foster and C. Kesselman (Eds.). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco (CA), 1999.

[23] M. S. Fox. An organizational view of distributed systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 11(1):70–80, January 1981.

[24] C. Handy. *Understanding Organizations*. Penguin Books, London (UK), 1976.

[25] F. Hattori, T. Ohguro, M. Yokoo, S. Matsubara, and S. Yoshida. Socialware: Multiagent systems for supporting network communities. *Communications of the ACM*, 42(3):55–61, March 1999.

[26] C. Iglesias, M. Garijo, and J. Gonzales. A survey of agent-oriented methodologies. In *Intelligents Agents IV: Agent Theories, Architectures, and Languages*, volume 1555 of *LNAI*, pages 317–330. Springer-Verlag, 1999.

[27] N. R. Jennings. An agent-based approach for building complex software systems. *Communications of the ACM*, 44(4):35–41, April 2001.

[28] T. Juan, A. Pierce, and L. Sterling. Roadmap: Extending the gaia methodology for complex open systems. In *Proceedings of the 1st ACM Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 3–10. ACM Press, Bologna (I), July 2002.

[29] E. A. Kendall. Agent software engineering with role modelling. In *Proceedings of the 1st International Workshop on Agent-Oriented Software Engineering*, volume 1957 of *LNCS*, pages 163–170. Springer Verlag, 2001.

[30] J. O. Kephart and D. M. Chess. The vision of autonomic computing. *IEEE Computer*, 36(1):41–50, January 2003.

[31] M. Kolp, P. Giorgini, and J. Mylopoulos. A goal-based organizational perspective on multi-agent architectures. In *Intelligent Agents VIII: Agent Theories, Architectures, and Languages*, volume 2333 of *LNAI*, pages 128–140. Springer Verlag, 2002.

[32] G. Leer and D. Hefferman. Expanding automotive electronic systems. *IEEE Computer*, 35(1):88–93, January 2002.

[33] J. Lind. *Iterative Software Engineering for Multiagent Systems: The MASSIVE Method*, volume 1994 of *LNCS*. Springer-Verlag, Berlin (D), 2001.

[34] P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7):30–40, July 1994.

[35] M. Mamei, F. Zambonelli, and L. Leonardi. Distributed motion coordination in co-fields. In *Proceedings of the 6th Symposium on Autonomous Decentralized Systems*, pages 63–70. IEEE CS Press, Pisa (I), April 2003.

[36] Z. Manna and A. Pnueli. *Temporal Verification of Reactive Systems — Safety*. Springer Verlag, Berlin (D), 1995.

[37] M. Mataric. Integration of representation into goal-driven behaviour-based robots. *IEEE Transactions on Robotics and Automation*, 8(3):59–69, December 1992.

[38] H. Mintzberg. *The Structuring of Organizations: A Synthesis of the Research*. Prentice Hall, Englewood Cliffs (NJ), 1979.

[39] Y. Moses and M. Tennenholtz. Artificial social systems. *Computers and Artificial Intelligence*, 14(3):533–562, 1995.

[40] J. Mylopoulos, L. Chung, and E. Yu. From object-oriented to goal-oriented requirements analysis. *Communications of the ACM*, 42(1):31–37, January 1999.

[41] P. Noriega. *Agent-mediated Auctions: The Fishmarket Metaphor*. Ph.D Thesis, Universitat Autonoma de Barcelona, Barcelona (E), 1997.

[42] J. Odell, H. Van Dyke Parunak, and C. Bock. Representing agent interaction protocols in uml. In *Proceedings of the 1st International Workshop on Agent-Oriented Software Engineering*, volume 1957 of *LNCS*, pages 121–140. Springer Verlag, 2001.

[43] V. Parunak. Go to the ant: Engineering principles from natural agent systems. *Annals of Operations Research*, 75:69–101, 1997.

[44] A. Ricci, A. Omicini, and E. Denti. Agent coordination infrastructures for virtual enterprises and workflow. *International Journal of Cooperative Information Systems*, 11(3):335–380, September 2002.

[45] M. Ripeani, A. Iamnitchi, and I. Foster. Mapping the gnutella network. *IEEE Internet Computing*, 6(1):50–57, January 2002.

[46] J. Rumbaugh, I. Jacobson, and G. Booch. *The Unified Modeling Language Reference Manual*. Addison-Wesley, Reading (MA), 1998.

[47] D. Schwabe, R. M. Guimaraer, and G. Rossi. Cohesive design of personalized web applications. *IEEE Internet Computing*, 6(2):34–43, April 2002.

[48] M. Shaw, R. DeLine, D. Klein, T. Ross, D. Young, and G. Zelesnik. Abstractions for software architecture and tools to support them. *IEEE Transactions on Software Engineering*, 21(4):314–335, April 1995.

[49] M. Shaw and D. Garlan. *Software Architectures: Perspectives on an Emerging Discipline*. Prentice Hall, Englewood Cliffs (NJ), 1996.

[50] O. Shehory and A. Sturm. Evaluation of modeling techniques for agent-based systems. In *Proceedings of the 5th International Conference on Autonomous Agents*, pages 624–631. ACM Press, Montreal (CA), June 2001.

[51] H. A. Simon. *Models of man*. Wiley, New York (NY), 1957.

[52] R.G. Smith. The contract net protocol: High level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, 29(12):1104–1113, December 1980.

[53] Y. Tahara, A. Ohsuga, and S. Honiden. Agent system development based on agent patterns. In *Proceedings of the 21st International Conference on Software Engineering*, pages 356–367. ACM Press, Los Angeles (CA), May 1999.

[54] David Tennenhouse. Embedding the Internet: Proactive computing. *Communications of the ACM*, 43(5):36–42, May 2000.

[55] M. Wood, S. A. DeLoach, and C. Sparkman. Multiagent system engineering. *International Journal of Software Engineering and Knowledge Engineering*, 11(3):231–258, April 2001.

[56] M. Wooldridge. Agent-based software engineering. *IEE Proceedings on Software Engineering*, 144(1):26–37, February 1997.

[57] M. Wooldridge. *An Introduction to Multiagent Systems*. John Wiley & Sons, Ltd., Indianapolis (IN), 2002.

[58] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.

[59] M. Wooldridge, N. R. Jennings, and D. Kinny. The Gaia methodology for agent-oriented analysis and design. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(3):285–312, 2000.

[60] M. J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, Cambridge (MA), 2000.

[61] F. Zambonelli, N. R. Jennings, A. Omicini, and M. Wooldridge. Agent-Oriented Software Engineering for Internet Applications. In *Coordination of Internet Agents: Models, Technologies, and Applications*, pages 326–346. Springer-Verlag, Berlin (D), 2001.

[62] F. Zambonelli, N. R. Jennings, and M. Wooldridge. Organizational rules as an abstraction for the analysis and design of multi-agent systems. *International Journal of Software and Knowledge Engineering*, 11(3):303–328, April 2001.

[63] F. Zambonelli and H. V. D. Parunak. Signs of a revolution in computer science and software engineering. In *3rd International Workshop on Engineering Societies in the Agents World*, volume 2577 of *LNCS*, pages 13–28. Springer Verlag, 2003.