Implementing a rule-driven approach for architectural self-configuration in collaborative activities using a graph rewriting formalism

Ismael Bouassida Rodriguez
LAAS-CNRS ; Université de Toulouse ; 7, avenue du Colonel Roche, F-31077 Toulouse, France
ibouassi@laas.fr

Karim Guennoun
Ecole Hassania des Travaux Publics
KM 7, Route D’EL JADIDA, Box 8108, Oasis, Casablanca, Morocco
guennoun@chtp-ac.ma

Khalil Drira
LAAS-CNRS ; Université de Toulouse ; 7, avenue du Colonel Roche, F-31077 Toulouse, France
khalil@laas.fr

Christophe Chassot
LAAS-CNRS ; Université de Toulouse ; INSA
7, avenue du Colonel Roche, F-31077 Toulouse, France
chassot@laas.fr

Mohamed Jmaiel
National Engineering School of Sfax, Redcad
Route de la Soukra Km 3.5 Box W 3038 Sfax, Tunisia
mohamed.jmaiel@enis.rnu.tn

ABSTRACT
Architectural adaptation is important for handling self-configuring properties of autonomic distributed systems. It can be achieved by model-based management of dynamic architectures. Describing dynamic architectures includes defining rules for reconfiguration management. Within this research context, several works have been conducted using formal specification to handle this complexity. Graph and graph rewriting-based approaches showed, through many studies, their appropriateness to tackle architectural adaptation problems. However, scalability of such approaches remains an open issue and has been rarely explored. In this paper, we investigate this issue. We introduce a graph-based general approach for handling of dynamic architectures, and we illustrate it within a scenario of collaboration support in Crisis Management Systems. We elaborate the formal models for dynamic architecture management. Using the French Grid GRID5000, we conducted an experimental study to assess the scalability of the elaborated models.

Categories and Subject Descriptors
D.2.11 [Software Engineering]: Software Architectures—Languages; D.2.9 [Software Engineering]: Management—Software configuration management; D.2.8 [Software Engineering]: Metrics—performance measures

General Terms
Design, Management

Keywords
Autonomic computing, Rule-driven, Graph rewriting

1. INTRODUCTION
The design of Self-configuring Systems that adapt their service to dynamically changing environments is among the main research directions of autonomic computing and communication [18]. Providing such solutions for distributed software systems supporting group communication requires dynamically managing evolving group membership and dynamically connecting deployment nodes. It also requires dynamically distributing software services and components on such remotely interconnected deployment nodes.

For a number of group communication-based applications, to anticipate reconfiguration is important. In order to be applicable in different situations, one has also to ensure tractability of the elaborated solutions when changing from some to thousands users, nodes, components and services.

Providing generic and scalable solutions for automated self-reconfiguration in group collaboration support systems can be driven by rule-based reconfiguration policies. We elaborate a graph-based modelling approach and structural models that may represent the different interaction dependencies from different configuration-related point of views: communication flows between the distributed machines, the networked deployment nodes, and the service composition.

Our solution is based on graph grammar rewriting. Architectural reconfiguration is handled by defining architectures as graphs where vertices correspond to deployment nodes, software services and their internal components. It provides graph transformation to specify rules for changing deployment architecture while being in conformance to an architectural style. Dealing with dynamically evolving architectures
requires at least describing the set of consistent architecture instances. This is mandatory to validate the management models and verify architectural constraint preservation. An architecture instance is considered consistent, if its corresponding description graph can be generated by a sequence of graph grammar productions. Our approach supports formal verification for correctness and safety proofs. We implemented a graph rewriting system that ensures automating our approach with a high performance making it tractable for large scale configurations.

To illustrate the proposed models and their transformations, we consider a case study of crisis management systems (CMS) involving several cooperating participants having different roles and functions. Related graph-based models are provided in section 4. Section 5 gives experimental results and analysis comments for scalability and performance of implemented models. Conclusion is provided in the last section.

2. RELATED WORK

Adaptation objectives, actions and properties are among the main facets of adaptability. They are studied and classified in this section.

2.1 Adaptability and adaptation

Adaptation is the operation to make changes to a program or an information system to maintain its functionalities and, if possible, to improve its performance in a certain execution environment. In the area of ubiquitous computing and context-awareness, adaptation is extended by the concept of adaptability that characterizes a system capability to change its behaviour to improve its performance or to continue its role in different environments. On the usefulness of adaptation, M. Satyanarayanan said in [36] “adaptability is necessary when there is a significant disparity between supply and demand of a resource.”

2.2 Goals

After developing an information system, many reasons can lead to adapt it. The reasons for this adaptation can be for corrective, evolutive or perfective purposes [22]. In some cases, we can notice that the application does not behave properly or as expected. The corrective adaptation is a solution to identify the application module that causes the problem and to replace it by a new correct module. This new module provides the same functionality as the former. For the evolutive adaptation, when developing an application, some features are not taken into account. With the changing needs of the user, the application must be extended with new features. This extension can be accomplished by adding one or more modules to provide new features or modifying existing modules to enrich their functionality while keeping the same application architecture. The objective of the perfective adaptation is to improve application performance. For example, we can have a module that receives a lot of requests and fails to meet them. To avoid system performance degradation, we can duplicate this module to share the requests with the existing one.

2.3 Classification

The adaptation solutions suggested in the literature can be classified as follow.

2.3.1 Design vs run-Time

Two different adaptability views may be distinguished: the design time adaptability [11, 16, 14] and the run time adaptability [7, 3].

For the first view, we can find design support tools handles the application development cycle and optimizes the resources for example. For the run time adaptability [17] presents several adaptation techniques which use proxy services, change model of interaction and reorganize application structure.

2.3.2 Local vs Distributed

Adaptation may have a local or a distributed scope. Adaptive components can be deployed on a single machine or distributed on several machines. In the first case, the adaptation is local and only local changes are performed. In the second case, it is distributed and synchronization problems between peer adaptive entities have to be managed [4].

2.3.3 Behavioural vs architectural

The adaptation solutions suggested in the literature distinguish behavioural and architectural aspects. The adaptation is behavioural (or algorithmic) when the behaviour of the adaptive service can be modified, without modifying its structure. Standard protocols such as TCP and specific protocols such as [39, 31] provide behaviour-based adaptation mechanisms. Behavioural adaptation is easy to implement but limits the adaptability properties.

The adaptation is architectural when the service composition can be modified [1, 13] dynamically. In self-adaptive applications components are created and connected, or removed and disconnected during the execution. The architectural changes respond to constraints related to the execution context involving, for example, variations of communication networks and processing resources. They may also respond to requirement evolution in the supported activities involving, for example, mobility of users and cooperation structure modification.

2.4 Objects of adaptation

Adaptation approaches target many levels of information systems: User interfaces, Content, Services, Middleware and Transport.

2.4.1 User interface adaptation

User interface adaptation consist in producing Human-Computer means that can be deployed and used on different types of terminals and which meet the user’s preferences. Major existing work in user interface adaptation is based on models that describe the different aspects of the interaction between humans and machines. These models are implemented in different XML or UML languages like UMLi [33] and XIML [34]. These models are used to produce the adequate user interface code corresponding to the given XML or UML description. In the existing user interface adaptations, we distinguish two techniques: User Interface transformations and User Interface generation. In the first technique, the adaptation process starts from a description language which is very close to the user interface code that must be generated. This solution was adopted by [29] to produce adapted Web pages to different terminals starting from an XML description. In this kind of adaptation, style sheets are used to specify replacement rules of XML tags by
scripts that can be directly used by the target device. The second approach consists in generating user interfaces code starting from a high level description which is completely independent from the target programming language of the user interface.

2.4.2 Content adaptation

Since the appearance of pervasive systems, the adaptation of multimedia content has been the subject of considerable research. Several techniques for adapting the delivered data to the user have been proposed. These techniques are based on textual transformations [27], image transcoding [38], or processing video and audio. One of the major issues in content adaptation is where the decision-making and transformations are made. In the literature, three general approaches have been proposed according to the location of adaptation processing between the source that hosts the content and destination that requests it: (i) on the content provider side, (ii) on the requester side, and (iii) at an intermediary (proxy) between the data source and the client. The content provider-side solutions have some drawbacks. Indeed, the changes made on the content induce a calculation load and consequent resource consumption on the server. However, this approach is very suitable for situations with low variability and low adaptation frequency regarding the simplicity of its implementation. But it is not reliable for cases where the adaptation is triggered frequently.

The content requester-side approach is suitable when the transmission characteristics are less critical than the display limits of the user device [28]. However, the usual complexity of adaptation processing hampers the adoption of this approach [32]. Indeed, the client’s terminal has generally very limited computing capacity, power and storage.

For the proxy solution, the flexibility of positioning the adaptation mechanisms on the content distribution point is a major advantage compared to the other approaches (provider and content sides). However, the proxy must be a trusted party by the provider and the requester of the content. In addition, the third party may charge for the service it provides and the resources it employs to perform the adaptation for the receiver. Therefore, accounting mechanisms should be incorporated in the proxy solution in order to keep track of the amount of resources utilized and the usage of the data.

2.4.3 Service adaptation

Service-Oriented Architecture (SOA) paradigm is based on dynamically publishing and discovering services. This kind of architectures provides the possibilities to dynamically compose services for adapting applications to contexts. Service descriptions are published, via the registry, by service providers and dynamically discovered by service requesters. There are various implementation technologies like Web Services or COM/DCOM (Component Object Model) [19] of Microsoft, the EJB (Enterprise Java Beans) [24] of Sun Microsystems or CCM (CORBA Component Model) [30] of OMG (Object Management Group).

2.4.4 Middleware adaptation

Other frameworks are proposed to provide adaptability for the middleware level. In [29], an adaptive framework for supporting multiple classes of multimedia services with different QoS requirements in wireless cellular networks is proposed. [37] proposes CME, a middleware architecture for service adaptation based on network awareness. CME is structured as the software platform both to provide network awareness to applications and to manage network resources in an adaptive fashion.

2.4.5 Transport adaptation

At the transport level, [15] provide frameworks for designing Transport protocols whose internal structure can be modified according to the application requirements and network constraints. Adaptation actions correspond to the replacement of a processing module or micro-protocol by another following a plug and play approach.

2.5 Model based adaptation approaches

There are many relevant contributions concerning system architecture adaptation. This kind of approaches uses model-based strategies to apply the necessary transformations on the systems architecture to adapt it to environment and requirement changes. These strategies define or reuse models describing the system software architectures. These models are also known as ADLs (Architecture description languages). We distinguish between these models: formal ADLs like graph grammars [20] and Petri nets [26], semantic ADLs using ontologies [40] and technical ADLs using XML deployment languages in general [10]. The technical ADLs can be proprietary or implementing the formal and the semantic architecture description models. These ADLs are used to guarantee the architecture evolving and correctness during the different predictable and unpredictable changes in the systems environment. The necessary actions to achieve such adaptations are specified in rules according to the application runtime context. [6] is an example of these approaches defining a complete model based architecture adaptation at the Service, content and user interface levels. [8] presents another model based method using graph grammars to adapt cooperative information systems to situation changes in the communication level.

Mobile networked systems for the support of group-wide activities have to provide dynamic adaptability for run-time and design-time changes. Such changes are induced by evolving requirements of the supported activities and by evolving resource constraints.

Designing and implementing self-adaptive communicating systems to support such emerging activities is complex. Mastering this complexity may be achieved by adopting model-based design approaches associated with automated management techniques for dynamic architectural and behavioral adaptability management.

The works exposed allow clarifying the different problems related to self-adaptability. The proposed solutions have different features related to the targeted goals (e.g. QoS, security...), the considered layers (e.g. Application layer, Transport layer...), the adaptation actions (e.g. rate control, congestion control...), the adaptation properties (e.g. architectural, behavioural...), and the way to manage the adaptation and its autonomy.

To face the different kinds of requirements and constraints evolving, behavioral and architectural adaptability is required at several levels simultaneously. This typical cross layering problematic requires managing coordination without which it can lead to performances way below the targeted ones.

For group communication-based cooperative activities, ab-
Struction levels have to be identified. Adaptation at the highest levels should be guided by the evolving of the activity requirements. Adaptation at the lowest levels should be driven by the changes due to device/network constraints.

Designing and implementing self-adaptive communicating systems is a complex task. To handle this complexity, several studies showed the need to lay on model-based design approaches associated with automated management techniques.

Static architectures are described by instances of components and interconnection links. The dynamic character of architectures requires additional description rules. Several works have addressed the dynamic architecture description, using different approaches [23, 2]. In order to guarantee the architecture evolving, correctness formal techniques are used. In particular, graphs represent a powerful expressive mean to specify respectively static and dynamic architecture aspects [25, 21]. For such approaches, graph vertices represent the software components, and the edges represent the links between these components. Dynamic architectures are described as graph grammars and architecture transformation is specified and ruled using graph rewriting models.

3. CASE STUDY

To expose the targeted problems and concepts, we consider the example of crisis management systems (CMS). In the sequel, we introduce this example and give two different execution steps and some related scenarios.

For CMS-like activities, collaboration is based on information exchange between mobile participants collaborating to achieve a common mission. A CMS team is composed of a number of participants having different roles: the controller, the mission coordinators, and field investigators. Each group of investigators is supervised by a coordinator (Figure 1). Each participant is associated with an identifier, a role and the devices he/she uses. Each participant performs different functions. The controller’s functions include monitoring and authorizing/managing actions to be done by coordinators and investigators. The controller is the entity which supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common mission. A CMS team is composed of a controller, who supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common mission. A CMS team is composed of a controller, who supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common mission. A CMS team is composed of a controller, who supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common mission. A CMS team is composed of a controller, who supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common mission. A CMS team is composed of a controller, who supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common mission. A CMS team is composed of a controller, who supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common mission. A CMS team is composed of a controller, who supervises the whole mission. The controller waits for data from his coordinators who synthesize the current situation of the mission. The controller has permanent energy resources and high communication and CPU capabilities. Coordinators and investigators supervise the mission. The coordinators have permanent energy resources from his coordinators who synthesize the current situation of the mission, coordinators, and field investigators. Each number of participants having different roles: The controller achieves a common

4. GRAPH-BASED MODELS FOR ARCHITECTURAL ADAPTATION

In this section, we introduce our general framework of models. We use models based on graph grammar and graph rewriting to handle the reconfiguration management.

Graph grammars constitute a powerful and very expressive formalism for style description. Moreover, theoretical work on this field provides formal means to specify and check constraints on these architectures [35]. Inspired from Chomsky’s generative grammars [9], graph grammars are defined, in general, as a classical system \( \langle AX; NT; T; P \rangle \), where \( AX \) is the axiom, \( NT \) is the set of the non-terminal vertices, \( T \) the set of terminal vertices, and \( P \) the set of transformation rules, also called grammar productions. An instance belonging to the graph grammar is a graph containing only terminal vertices and is obtained starting from axiom \( AX \) by applying a sequence of productions in \( P \). There are different approaches for the definition of graph grammar production structure and the mechanisms used for the specification of their execution. In our approach, a graph grammar is of the form \( \langle AX; NT; T; P \rangle \) where \( AX \), \( NT \) and \( T \) keep the same significance as before. We use productions of type \( \langle L; K; R; C \rangle \) where \( \langle L; K; R \rangle \) corresponds to the structure of

![Figure 1: CMS architecture](image)
a Double PushOut (DPO) production [12] and where \( C \) is a set of connection instructions. The instructions belonging to \( C \) are of the edNCE type [35]. They are specified by a system \((n, s, d, d')\) where \( n \) corresponds to a vertex belonging to the daughter graph \( R \), \( p \) and \( q \) are two edge labels, \( \delta \) is a vertex label, and \( d \) and \( d' \) are elements of the set in, out. For example, a production defined by the system \((L; K; R; (n, s, d, d'))\) is applicable to a graph \( G \) if it contains an occurrence of the mother graph \( L \). The application of this production involves transforming \( G \) by deleting the subgraph \((\text{Del} = L \setminus K)\) and adding the subgraph \((\text{Add} = R \setminus K)\) while the subgraph \( K \) remains unchanged. All dangling edges will be removed.

The execution of the connection instruction implies the introduction of an edge between the vertex \( n \) belonging to the daughter graph \( R \) and all vertices \( n' \) that are \( p \)-neighbours\(^1\) of and \( d \)-neighbours\(^2\) of \( n \). This edge is introduced following the direction indicated by \( d' \) and labelled by \( q \).

An example is given in Figure 2. Production \( p3 \) has the same structure and transformation logic as in the DPO approach: node 5 is not removed even if it is matched with node 2 belonging to the \( L \) pattern because it also belongs to the \( K \) pattern. The example also considers the edNCE connection instructions \( c1 \), \( c2 \) and \( c3 \) allowing the correct addition of edges connecting nodes 2 and 6 and connecting nodes 3 and 7.

Following the commonly used conventions, we consider that vertices represent communicating entities (e.g., services, components) and edges correspond to their related interdependencies (e.g., communication links, composition dependencies).

An example of an architecture instance is given in figure 3. This instance includes, in addition to the controller (Cont), two coordinators (coord) that manage respectively three and two investigators (Inv). The graph edges are labelled by the exchanged data types \((D/P)\). Each participant have two attributes: the identifier and the deployment host. For example an investigator node is represented by a label like that \( \text{inv(identifier, deploymenthost)} \). Architectural reconfiguration management consisting in transforming the system architecture evolving situation is formally ruled using graph rewriting techniques.

\(^1\)\(p\)-neighbours of a vertex \( n \) are all vertices \( n' \) such that there exists an edge labelled by \( p \) which connects \( n \) and \( n' \).

\(^2\)\(d\)-neighbours if \( d = \text{in} \) and \( d \)-neighbours otherwise.

5. THE GRAPH GRAMMAR FOR ARCHITECTURE RECONFIGURATION

In the following, we provide an example of graph grammar for our case study to implement architecture reconfiguration.

5.1 Managing changes in activity organisation

The presented rewriting rule allows transforming the architecture from for moving the exploration step to the action step. The graph grammar is reduced to a single production grammar \( P_{\text{exp} \rightarrow \text{act}} \) (table1) which is parameterized by the investigator identifier (here, noted A) who has discovered the critical situation. The architecture is transformed by splitting the communication channels between the coordinator and the other investigators into a communication channel of type \( P \), between these investigators and A into another communication channel of type \( D \). Figure 4 gives an example of the application of this rule to move the architecture from the exploration step to the action step. Investigator 2 plays the role of the critical investigator A. \( P_{\text{exp} \rightarrow \text{act}} \) allows the correct generation of the communication channels. The connection instruction \( ic1 \) allows to change the labels of the coordinator to the investigators from \(< D, P >\) to \(< P >\). The connection instruction \( ic2 \) allows to connect the investigator A to the others investigator by an edge labelled \(< P >\).

5.2 Managing changes in group membership

We elaborate a graph grammar for our case study to manage the group extension. In the exploration step, the investigators produce D-data continuously and P-data periodically. The coordinators send P-data periodically to the controller. In the graph grammar \( GG \) (table2) we have four productions.
The productions $p_1$, $p_2$ and $p_3$ allow us to add new participants to the mission. The production $p_1$ allows to initiate the group by connection an investigator, a coordinator and a controller. The production $p_2$ allows to connect a coordinator and the edges necessary to that. According to the description of the case study, in the group, it is not possible to have a coordinator without an investigator. So the production $p_2$ adds an investigator to the added coordinator. The production $p_3$ allows to add an investigator to an existent coordinator. Such a propriety constitute the style of the architecture. The production $p_4$ allows to finish the generation process and eliminate the non-terminal $Temp$. The example shown in figure 3, can be generated\(^3\) by this sequence: $p_1,p_2,p_3,p_3,p_4$. The existence of this generation path prove the consistency of this architecture instance.

6. EXPERIMENTAL STUDY

We conducted an experimental study of our graph rewriting system using the rules of the architecture adaptation models presented above.

These experiments have been achieved under grid GRID5000 [5], an experimental grid platform that interconnects several clusters geographically distributed in France. Considered grid nodes are single core machines (Sun Fire V20z) with an AMD Opteron 248 processor with 2.2 GHz of CPU speed and 2 GB of RAM. The provided network bandwidth is up to 10GB/s. We are studying scalability and limits of this model to investigate the tractability of an anticipation approach for adaptability.

The results, presented here, were obtained for graph grammar $GG$ of section 4 and the reconfiguration rules executed to handle critical situation discovery (grammar production $P_{\text{act}}$ in table 1 of section 4). The experimental performance evaluation focuses execution time.

6.1 Experimentations for adapting architecture to organization changes

In this section we present, the results related to the execution of the $P_{\text{act}}$ rewriting rule. We are studying scalability and limits of our model-driven approach implementation.

We study the execution of the experiment (figure 5), when the architecture description graph size growth. This corresponds to the application of production $p_1$ of $P_{\text{act}}$ (Table 1). The implementation performs well and execution time is under 0.3 second for a 10 000 vertex graph (i.e. a system composed of 10 000 participants).

6.2 Experimentation adapting architecture to group membership changes

We generate the first N members of this graph grammar corresponding to the N first consistent instances of the ar-

---

\(^3\)This is not the unique generation path.
architecture. This corresponds to applying productions $p_1$, $p_2$, and $p_3$ of $GG$ (Table 2) whenever it is possible to obtain all consistent architecture graphs. The application of production $p_4$ ensures the elimination of the non-terminal nodes in the graphs. The production applications include finding all graph matchings between the rules graph and transforming the architecture graph by adding nodes and edges. The production $p_2$ adds two nodes and two edges. The production $p_1$ adds one node and one edge.

![Figure 6](image)

**Figure 6:** Execution time of applying $p_1$, $p_2$, $p_3$, and $p_4$ of $GG$ wrt graph number

The system, takes about nine hours to calculate the 400 000 first consistent instances and more than 1 500 000 consistent instances are generated in about 194 hours. These values remains acceptable for design-time purposes. Using a unique grid node, the system generates almost the 400 000 first consistent instances in about 140 hours.

7. CONCLUSION

In this paper, we have studied a graph-based approach for implementing a model-based management of architectural reconfiguration. Our approach is useful when designing self-configuring software for large scale activities support. It supports scalable management of automated reconfiguration while allowing formal proofs to be developed for ensuring the consistency of the generated configurations. The models presented in this paper have been elaborated for the design of self-configuring Crisis Management Systems. The scalability study provided in this paper has been conducted to assess the tractability of our approach when dealing with large scale group applications.

8. REFERENCES


