

₁ A Unified Modeling Framework to Abstract
₂ Knowledge of Dynamically Adaptive Systems

₃ Ludovic Mouline

₄ May 1, 2019

Abstract

Vision: As state-of-the-art techniques fail to model efficiently the evolution and the uncertainty existing in dynamically adaptive systems, the adaptation process makes suboptimal decisions. To tackle this challenge, modern modeling frameworks should efficiently encapsulate time and uncertainty as first-class concepts.

Context Smart grid approach introduces information and communication technologies into traditional power grid to cope with new challenges of electricity distribution. Among them, one challenge is the resiliency of the grid: how to automatically recover from any incident such as overload? These systems therefore need a deep understanding of the ongoing situation which enables reasoning tasks for healing operations. **Abstraction** is a key technique that provided an illuminating description of systems, their behaviors, and/or their environments alleviating their complexity. **Adaptation** is a cornerstone feature that enables reconfiguration at runtime for optimizing software to the current and/or future situation.

Abstraction technique is pushed to its paramountcy by the model-driven engineering (MDE) methodology. However, information concerning the grid, such as loads, is not always known with absolute confidence. Through the thesis, this lack of confidence about data is referred to as **data uncertainty**. They are approximated from the measured consumption and the grid topology. This topology is inferred from fuse states, which are set by technicians after their services on the grid. As humans are not error-free, the topology is therefore not known with absolute confidence. This data uncertainty is propagated to the load through the computation made. If it is neither present in the model nor not considered by the adaptation process, then the adaptation

1 process may make suboptimal reconfiguration decision.

2 The literature refers to systems which provide adaptation capabilities as dynamically
3 adaptive systems (DAS). One challenge in the grid is the phase difference between the
4 monitoring frequency and the time for actions to have measurable effects. Action with
5 no immediate measurable effects are named **delayed action**. On the one hand, an
6 incident should be detected in the next minutes. On the other hand, a reconfiguration
7 action can take up to several hours. For example, when a tree falls on a cable and cuts
8 it during a storm, the grid manager should be noticed in real time. The reconfiguration
9 of the grid, to reconnect as many people as possible before replacing the cable, is done
10 by technicians who need to use their cars to go on the reconfiguration places. In a fully
11 autonomous adaptive system, the reasoning process should be considered the ongoing
12 actions to avoid repeating decisions.

13 *Problematic* **Data uncertainty and delayed actions are not specific to smart**
14 **grids.**

15 First, data are, almost by definition, uncertain and developers always work with
16 estimates. Hardware sensors have by construction a precision that can vary accord-
17 ing to the current environment in which they are deployed. A simple example is the
18 temperature sensor that provides a temperature with precision to the nearest degree.
19 Software sensors approximate also values from these physical sensors, which increases
20 the uncertainty. For example, CPU usage is computed counting the cycle used by a
21 program. As stated by Intel, this counter is not error-prone¹.

22 Second, it always exists a delay between the moment where a suboptimal state is
23 detected by the adaptation process and the moment where the effects of decisions taken
24 are measured. This delayed is due to the time needed by a computer to process data
25 and, eventually, to send orders or data through networks. For example, migrating a
26 virtual machine from a server to another one can take several minutes.

27 **Through this thesis, I argue that this data uncertainty and this delay**
28 **cannot be ignored for all dynamic adaptive systems.** To know if the data un-
29 certainty should be considered, stakeholders should wonder **if this data uncertainty**

¹<https://software.intel.com/en-us/itc-user-and-reference-guide-cpu-cycle-counter>

1 **affects the result of their reasoning process, like adaptation.** Regarding delayed
2 action, they should verify **if the frequency of the monitoring stage is lower than**
3 **the time of action effects to be measurable.** These characteristics are common
4 to smart grids, cloud infrastructure or cyber-physical systems in general.

5 *Challenge* These problematics come with different challenges concerning the represen-
6 tation of the knowledge for DAS. The global challenge address by this thesis is: **how**
7 **to represent the uncertain knowledge allowing to efficiently query it and to**
8 **represent ongoing actions in order to improve adaptation processes?**

9 *Vision* **This thesis defends the need for a unified modeling framework which**
10 **includes, despite all traditional elements, temporal and uncertainty as first-**
11 **class concepts.** Therefore, a developer will be able to abstract information related to
12 the adaptation process, the environment as well as the system itself.

13 Concerning the adaptation process, the framework should enable abstraction of the
14 actions, their context, their impact, and the specification of this process (requirements
15 and constraints). It should also enable the abstraction of the system environment and its
16 behavior. Finally, the framework should represent the structure, behavior and specifi-
17 cation of the system itself as well as the actuators and sensors. All these representations
18 should integrate the data uncertainty existing.

19 *Contributions* Towards this vision, this document presents two approaches: a temporal
20 context model and a language for uncertain data.

21 The temporal context model allows abstracting past, ongoing and future actions
22 with their impacts and context. First, a developer can use this model to know what the
23 ongoing actions, with their expect future impacts on the system, are. Second, she/he
24 can navigate through past decisions to understand why they have been made when they
25 have led to a sub-optimal state.

26 The language, named Ain'tea, integrates data uncertainty as a first-class concept. It
27 allows developers to attach data with a probability distribution which represents their
28 uncertainty. Plus, it mapped all arithmetic and boolean operators to uncertainty prop-
29 agation operations. And so, developers will automatically propagate the uncertainty

1 of data without additional effort, compared to an algorithm which manipulates certain
2 data.

3 *Validation* Each contribution has been evaluated separately. The language has been
4 evaluated through two axes: its ability to detect errors at development time and its
5 expressiveness. Ain'tea can detect errors in the combination of uncertain data earlier
6 than state-of-the-art approaches. The language is also as expressive as current ap-
7 proaches found in the literature. Moreover, we use this language to implement the load
8 approximation of a smart grid furnished by an industrial partner, Creos S.A.².

9 The context model has been evaluated through the performance axis. The disser-
10 tation shows that it can be used to represent the Luxembourg smart grid. The model
11 also provides an API which enables the execution of query for diagnosis purpose. In
12 order to show the feasibility of the solution, it has also been applied to the use case
13 provided by the industrial partner.

14 **Keywords:** dynamically adaptive systems, knowledge representation, model-driven
15 engineering, uncertainty modeling, time modeling

²Creos S.A. is the power grid manager of Luxembourg. <https://www.creos-net.lu>

1 Table of Contents

2	1 Introduction	1
3	1.1 Introduction	2
4	1.2 Use case: Luxembourg smart grid	2
5	1.3 General background	2
6	2 TKM: a temporal knowledge model	3
7	2.1 Introduction	5
8	2.1.1 Motivation	6
9	2.1.2 Background	12
10	2.1.3 Use case scenario	15
11	2.2 Knowledge formalization	16
12	2.2.1 Formalization of the temporal axis	17
13	2.2.2 Formalism	18
14	2.2.3 Application on the use case	22
15	2.3 Actions as context information	27
16	2.3.1 Example: cluster management	27
17	2.3.2 Action planning	28
18	2.3.3 Linking action planning and context representation	29
19	2.3.4 Enabled reasoning	30
20	2.4 Modeling the knowledge	34
21	2.4.1 Parent element: <i>TimedElement</i> class	34
22	2.4.2 Knowledge metamodel	35

1	2.4.3	Context metamodel	36
2	2.4.4	Requirement metamodel	37
3	2.4.5	Action metamodel	37
4	2.5	Validation	39
5	2.5.1	Validation of action consideration	39
6	2.5.2	Validation for the temporal graph	43
7	2.6	Threat to validity	51
8	2.7	Conclusion	51
9	Abbreviations		i
10	Glossary		iii

1

2 Introduction

3 Contents

4	1.1 Introduction	2
5		
6	1.2 Use case: Luxembourg smart grid	2
7		
8	1.3 General background	2
10		

11 *Model-driven engineering methodology and dynamically adaptive systems approach*
12 *are combined to tackle new challenges brought by systems nowadays. After introducing*
13 *these two software engineering techniques, I give one example of such systems: the*
14 *Luxembourg smart grid. I will also use this example to highlight two of the problematics:*
15 *uncertainty of data and delays in actions. Among the different challenges which are*
16 *implied by them, I present the global one addressed by the vision defended in this thesis:*
17 *modeling of temporal and uncertain data. This global challenge can be addressed by*
18 *splitting up in several ones. I present two of them, which are directly tackled by two*
19 *contributions presented in this thesis.*

1 Introduction

2 Use case: Luxembourg smart grid

3 Should contain: - veg iqu grqaub

4 General background

5 should contain: - MDE / metamodel / model - DAS

1

2 A temporal knowledge meta-model to represent, 3 reason and diagnose decisions, their circumstances 4 and their impacts

5

Contents

6	2.1 Introduction	5
7		
8	2.2 Knowledge formalization	16
9	2.3 Actions as context information	27
10	2.4 Modeling the knowledge	34
11	2.5 Validation	39
12	2.6 Threat to validity	51
13	2.7 Conclusion	51
14		
15		
16		

17 *Adaptation processes are executed with a high frequency to react to any incident*
 18 *whereas the delay for decision applications are constrained by the time to execute the*
 19 *delayed actions. We identified two problems that result from these different paces. First,*
 20 *not considered unfinished actions, together with their expected effects, over time lead*
 21 *upcoming analysis phases potentially make suboptimal decisions. Second, explanations*
 22 *of adaptation processes remain challenging due to the lack of tracing ability of current*
 23 *approaches. To tackle this problem, we first propose a knowledge formalism to define*
 24 *the concept of a decision. Second, we describe a novel temporal knowledge model to*

1 *represent, store and query decisions as well as their relationship with the knowledge*
2 *(context, requirements, and actions). We validate our approach through a use case*
3 *based on the smart grid at Luxembourg. We also demonstrate its scalability both in*
4 *terms of execution time and consumed memory.*

Introduction

TODO: We consider that decision, delayed action, context and knowledge have been defined in the global introduction

Adaptive systems have proven their suitability to handle the increasing complexity of systems and their ever-changing environment. To do so, they make adaptation decisions, in the form of actions, based on high-level policies. For instance, the OpenStack Watcher project [OpenStack:Watcher:Wiki] implements the MAPE-k loop to assist cloud administrators in their activities to tune and rebalance their cloud resources according to some optimization goals (e.g., CPU and network bandwidth). For readability purpose, we refer to adaptation decision as decision in the remaining part of this document.

Despite the reactivity of adaptation processes, impacts of their decisions can be measurable long after they have been taken. We identified two problematics caused by this difference of paces:

- How to enable reasoning over unfinished actions and their expected effects?
- How to diagnose the self-adaptation process?

To address them, we propose a temporal knowledge model which can trace decisions over time, along with their circumstances and effects. By storing them, the adaptation process could consider ongoing actions with their expected effects, also called impacts. Plus, in case of faulty decisions, developers may trace back their effects to their circumstances. Our current approach is limited to the representation of measurable effects of any decision, and therefore action.

The meta-model allow structuring and storing the state and behavior of a running adaptive system, together with a high-level API to efficiently perform diagnosis routines. Our framework relies on a temporal model-based solution that efficiently abstracts decisions and their corresponding circumstances. Specifically, based on existing approaches for modeling and monitoring adaptation processes, we identify a set of properties that characterize context, requirements, and actions in self-adaptive systems. Then, we formalize the common core concepts implied in adaptation processes, also referred to as knowledge, by means of temporal graphs and a set of relations that trace decisions

1 impact to circumstances. Finally, thanks to exposing common interfaces in adaptive
2 processes, existing approaches in requirements and goal modeling engineering can be
3 easily integrated into our framework.

4 The rest of this chapter is structured as follows. In the remaining part of this section,
5 we motivate our approach, we summarize core concepts manipulated in adaptation
6 processes, and we present a use case scenario based on the Luxembourg Smart Grid
7 (*cf.* Chapter *TODO: add ref*). *TODO: Update with last version* Then, we
8 provide a formal definition of these concepts in Section 2.2. Later, we describe the
9 proposed data model in Section 2.4. In Section 2.5, we demonstrate the applicability
10 of our approach by applying it to the smart grid example. We conclude this chapter in
11 Section 2.7.

12 Motivation

13 Delayed action

14 In this section, we motivate the need to reason over delayed actions. To do so, we
15 first give four examples of these actions. Then we detail why the effects of actions
16 should be considered. Finally, we summarize and motivate the need for incorporating
17 actions and their effects on the knowledge.

18 **Delayed action examples** Until here, we have claimed that adaptation processes
19 should handle delayed actions. In order to show their existence, we give four different
20 examples: two based on our use case, one on cloud infrastructure and a last one on
21 smart homes. From our understanding, three phenomena can explain this delay: the
22 time to execute an action(s) (Example 1), the time for the system to handle the new
23 configuration (Example 3) and the inertia of the measured element (Example 2 and 4).

24 **Example 1: Modification of fuse states in smart grids** Even if the Luxem-
25 bourg power grid is moving to an autonomous one, not all the elements can be remotely
26 controlled. One example is fuses that still need to be open or closed by a human. Open
27 and close actions in the Luxembourg smart grid both imply technicians who are con-
28 tacted, drive to fuse places and manually change their states. If several fuses need to
29 be changed due to one decision, only one technician will drive to them, sequentially,

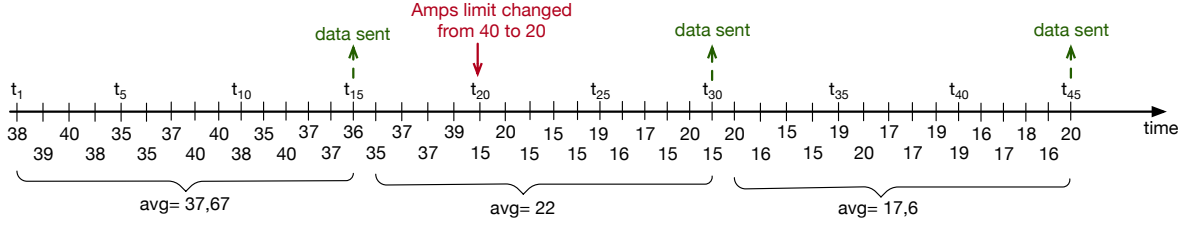


Figure 2.1: Example of consumption measurement before and after a limitation of amps has been executed at t_{20} .

1 and executes the modifications. For example, in our case, our industrial partner asks
2 us to consider that each fuse modification takes 15 min whereas any incident should
3 be detected in the minute. Let's imagine that an incident is detected at 4 p.m. and
4 can be solved by modifying three fuses. The incidents will be seen as resolved by the
5 adaptation process at 4 p.m. + 15 min * 3 = 4:45 p.m. In this case, the delay of the
6 action is due to the execution time that is not immediate.

7 **Example 2: Reduction of amps limit in smart grids¹** In its smart grid
8 project, Creos S.A. envisages controlling remotely amps limits of customers. Customers
9 will have two limits: a fixed one, set at the beginning, and a flexible one, remotely
10 managed. The action to remotely change amps limits will be performed through specific
11 plugs, such as one for electric vehicles. Even if the action is near instant, due to
12 how power consumption is collected, its impacts would not be visible immediately.
13 Indeed, data received by Creos S.A. corresponds to the total energy consumed since the
14 installation. From this information, only the average of consumed data for the last
15 period can be computed.

16 In Figure 2.1, we depict a scenario that shows the delay between the action is
17 executed and the impacts are measured. Each time point represents one minute, with
18 the consumption at this moment.

19 Let's imagine a customer who has his or her limit set to 40 amps² and consumes
20 near this limit. We consider that data are sent every 15 min. After receiving data

¹This example is based on randomly generated data. As this action is not yet available on the Luxembourg smart grid, we miss real data. However, it reflects an hypothesis shared with our partner.

²The user cannot consume more than 40 amps at a precise time t_i .

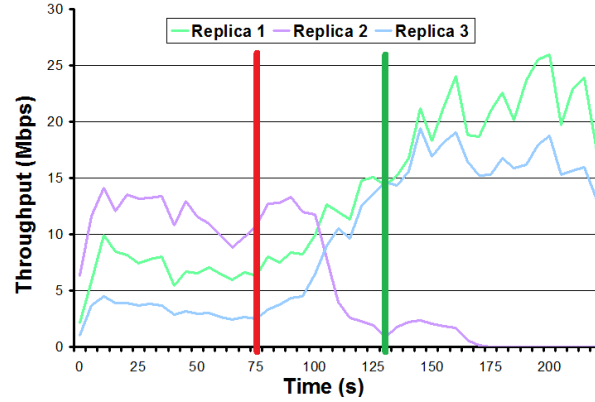


Figure 2.2: Figure extracted from [DBLP:conf/nsdi/WangBR11]. The red bar depicted the moment when Replica 2 stop receiving new connections. The green one represents the moment where all the rules in the load balancer stop considering R2. Despite these two actions, the throughput of the machine does not drop to 0 due to existing and active connections.

1 sent t_{15} and processing them, the adaptation process detects an overload and decides to
2 reduce the limits to 20 amps for the customer. However, considering the delay for data
3 to be collected and the one to send data³, the action is received and executed at t_{20} . At
4 t_{30} , new consumption data is sent, here equals 22 amps. Here, there are two situations.
5 First, this reduction was enough to fix the overload. Even in this idealistic scenario, the
6 adaptation process must wait at worst 15 min ($t_{30} - t_{15}$) to see the resolution (without
7 considering the communication time). Second, this reduction was not enough - as the
8 adaptation process considered that the consumption data will be at worst 20 amps and
9 here it is 22. Before seeing the incident as solved and knowing that the decision fixed
10 the incident, the adaptation process should wait for new data, sent at t_{45} , *i.e.*, around
11 30 min ($t_{45} - t_{15}$) after the detection.

12 In this case, the delay of this action can be explained by the inertia in the average
13 of the consumption.

14 **Example 3: Switching off a machine from a load balancer** An exam-
15 ple based on cloud infrastructure of delayed actions is to remove a machine from a

³Reminder: the smart grid is not built upon a fast network such a fiber network.

1 load balancer, for example during a scale down operation. Scale down operations
2 allows cloud managers to reduce allocated resources for a specific task. It is used
3 either to reduce the cost of the infrastructure or to reallocate them to other tasks.
4 In [DBLP:conf/nsdi/WangBR11], Wang *et al.*, present a load-balancing algorithm.
5 In their evaluation, they present the figure depicted in Figure 2.2 that shows the evo-
6 lution of the throughput after the server Replica 2 (R2) is removing from the load
7 balancer. The red bar shows the moment where R2 stop receiving new connection and
8 the green the moment where it is removed from the load balancer algorithm. However,
9 despite these actions have been taken, R2 should finish the ongoing tasks that it is
10 executing. This explains why the throughput is progressively decreasing to 0 and there
11 is a delay of around 100s between the red bars and the moment where R2 stop being
12 active.

13 This example shows a delayed action due to the time required by the system to
14 handle the new configuration.

15 **Example 4: Modifying home temperature through a smart home system**
16 Smart home systems have been implemented in order to manage remotely a house or to
17 perform automatically routines. For example, it allows users to close or open blinds from
18 their smartphones. Based on instruction temperatures, smart home systems manage the
19 heating or cooling system to reach them at the desired time. However, heating or cooling
20 a house is not immediate, it can take several hours before the targeted temperature is
21 reached. Plus, if the temperature sensor and the heating or cooling system are not
22 placed nearby, the new temperature can take time before being measured. This can
23 be explained due to the temperature inertia plus the delay for the temperature to be
24 propagated.

25 Through these four examples, we show that delayed actions can be found in different
26 kinds of systems, from CPS to cloud infrastructure. However, not only knowing that
27 an action is running is important but also knowing its expecting effect. We detail this
28 point in the following section.

29 **The need to consider effects** In the previous section, we show the existence of
30 delayed actions. One may argue that action statuses are already integrated into the

1 knowledge. For example, the OpenStack Watcher framework stores them in a database⁴,
2 accessible through an API. However, for the best of our knowledge Watcher does not
3 store the expecting effects of each action. While the adaptation process knows what
4 action is running, it does not know what it should expect from them.

5 Considering our example based on the modification of fuses, if the system knows
6 that the technician is modifying fuse states, it does not know what would be the effects.
7 In this case, when the adaptation process analyzes the system context it may wonder:
8 what will be the next grid configuration? How the load will be balanced? Will the
9 future configuration fix all the current incidents? If the effects are not considered by
10 the adaptation process, then it may take suboptimal decisions.

11 Let's exemplify this claim through a scenario based on the fuse example (*cf.* Exam-
12 ple 1). As explained before, the overload detected at 4 p.m. takes around 45 min to be
13 fixed. The system marks this incident as "being resolved". In addition to this informa-
14 tion, the knowledge contains another one saying that it is being solved by modifying
15 three fuses. However, during the resolution stage, a cable is also being overloaded. The
16 adaptation process has two solutions. It can either wait for the end of the resolution
17 of the first incident to see if both overloaded elements will be fixed or it takes other
18 actions without considering the ongoing actions and their impacts. Applying the first
19 strategy may make the resolution of the second incident late, whereas the second one
20 may generate a suboptimal sequence of actions. For example, the second modifications
21 may undo what has been done before or both actions may be conflicting.

22 **Conclusion** Actions, like fuse modification in a smart grid or removing a server
23 from a load balancer, generated during by adaptation processes could take time upon
24 completion. Moreover, the expected effects resulting from such action is reflected in the
25 context representation only after a certain delay. One used workaround is the selection,
26 often empirically, of an optimistic time interval between two iterations of the MAPE-K
27 loop such that this interval is bigger than the longest action execution time. However,
28 the time to execute an action is highly influenced by system overload or failures, making
29 such empirical tuning barely reliable. We argue that by enriching context representation

⁴<https://docs.openstack.org/watcher/latest/glossary.html#watcher-database-definition>

1 with support for past and future planned actions and their expected effects over time,
2 we can highly enhance reasoning processes and avoid empirical tuning.

3 The research question that motivates our work is thus: how to enable reasoning over
4 unfinished actions and their expected effects?

5 Fined and rich context information directly influences the accuracy of the actions
6 taken. Various techniques to represent context information have been proposed; among
7 which we find the models@run.time [DBLP:journals/computer/MorinBJFS09;
8 DBLP:journals/computer/BlairBF09]. The models@run.time paradigm inherits
9 model-driven engineering concepts to extend the use of models not only at design time
10 but also at runtime. This model-based representation has proven its ability to structure
11 complex systems and synthesize its internal state as well as its surrounding environment.

12 **Diagnosis support**

13 Faced with growingly complex and large-scale software systems (e.g. smart grid
14 systems), we can all agree that the presence of residual defects becomes unavoid-
15 able [DBLP:conf/icse/BarbosaLMJ17; DBLP:conf/icse/MongielloPS15; DBLP:conf/icse/H.
16 Even with a meticulous verification or validation process, it is very likely to run into
17 an unexpected behavior that was not foreseen at design time. Alone, existing formal
18 modeling and verification approaches may not be sufficient to anticipate these fail-
19 ures [DBLP:conf/icse/TaharaOH17]. As such, complementary techniques need to
20 be proposed to locate the anomalous behavior and its origin in order to handle it in a
21 safe way.

22 As there might be many probable causes behind an abnormal behavior, developers
23 usually perform a set of diagnosis routines to narrow down the scope or origin of the fail-
24 ure. One way to do so is by investigating the satisfaction of its requirements and the de-
25 cisions that led to this system state, as well as their timing [DBLP:conf/iceccs/BencomoWSW12].
26 In this perspective, developers may set up a set of systematic questions that would help
27 them understand why and how the system is behaving in such a way. These questions
28 may comprise:

- 29 • what goal(s) the system was trying to reach by executing a tactic a ?
- 30 • what were the circumstances used by a decision d and its expected impact on the

context?

- what decision(s) influenced the system’s context at a time t ?

Bencomo *et al.*, [DBLP:conf/iceccs/BencomoWSW12] argue that comprehensive explanation about the system behavior contributes drastically to the quality of the diagnosis, and eases the task of troubleshooting the system behavior. To enable this, we believe that adaptive software systems should be equipped with traceability management facilities to link the decisions made to their **(i) circumstances, that is to say, the history of the system states and the targeted requirements, and (ii) the performed actions with their impact(s) on the system.** In particular, an **adaptive system should keep a trace of the relevant historical events.** Additionally, it should be able to **trace the goals intended to be achieved by the system to the adaptations and the decisions that have been made, and vice versa.** Finally, in order to enable developers to interact with the system in a clear and understandable way, appropriate abstraction to **enable the navigation of the traces and their history should also be provided.** Unfortunately, suitable solutions to support these features are under-investigated.

Existing approaches [hassel13; heinrich14; ehlers11; DBLP:conf/icse/MendoncaAR14; DBLP:conf/icse/CasanovaGSA14; DBLP:conf/icse/IftikharW14a] are accompanied by built-in monitoring rules and do not allow to interact with the underlying system in a simple way. Moreover, they do not keep track of historical changes as well as causal relationships linking requirements to their corresponding adaptations. Only flat execution logs are stored.

Background

Before formalizing and modeling decisions and their circumstances, we abstract common concepts implied in an adaptation process. We refer to these concepts as the knowledge.

General concepts of adaptation process

Similar to the definition provided by Kephart [DBLP:journals/computer/KephartC03], IBM defines adaptive systems as “a computing environment with the ability to manage

1 itself and **dynamically adapt** to change in accordance with **business policies and**
2 **objectives**. [These systems] can perform such activities based on **situations they**
3 **observe or sense in the IT environment [...]** [computing2006architectural].

4 Based on this definition, we can identify three principal concepts involved in adap-
5 tation processes. The first concept is *actions*. They are executed in order to perform
6 a dynamic adaptation through actuators. The second concept is **business policies**
7 **and objectives**, which is also referred to as the **system requirements** in the domain
8 of (self-)adaptive systems. The last concept is the observed or sensed **situation**, also
9 known as the **context**. The following subsections provide more details about these
10 concepts.

11 Context

12 In this thesis, we use the widely accepted definition of context provided by Dey [DBLP:journals/puc
13 “Context is **any information that can be used to characterize** the situation
14 of an entity. An entity is a person, place, or object that is considered relevant to
15 the interaction between a user and [the system], including the user and [the sys-
16 tem] themselves”. In this section, we list the characteristics of this information based
17 on several works found in the literature [DBLP:conf/pervasive/HenricksenIR02;
18 chong2007context; DBLP:conf/seke/0001FNMKT14; bettini2010survey; DBLP:journals/co
19 We use them to drive our design choices of our Knowledge meta-model (cf. Sec-
20 tion *TODO: Add ref*).

21 **Volatility** Data can be either **static** or **dynamic**. Static data, also called frozen, are
22 data that will not be modified, over time, after their creation [DBLP:conf/pervasive/HenricksenIR0
23 DBLP:journals/comsur/MakrisSS13; bettini2010survey; chong2007context].
24 For example, the location of a machine, the first name or birth date of a user can be
25 identified as static data. Dynamic data, also referred to as volatile data, are data that
26 will be modified over time.

27 **Temporality** In dynamic data, sometimes we may be interested not only in stor-
28 ing the latest value, but also the previous ones [DBLP:conf/seke/0001FNMKT14;
29 DBLP:conf/pervasive/HenricksenIR02; chong2007context]. We refer to these
30 data as **historical** data. Temporal data is not only about past values, but also future

ones. Two kinds of future values can be identified, **predicted** and **planned**. Thanks to machine learning or statistical methods, dynamic data values can be **predicted**. **Planned** data are set by a system or a human to specify planned modification on the data.

Uncertainty One of the recurrent problems facing context-aware applications is the data uncertainty [DBLP:conf/dagstuhl/LemosGMSALSTVWBBBBBCDDEGGGGIKKLMM; DBLP:conf/pervasive/HenricksenIR02; DBLP:journals/comsur/MakrisSS13; bettini2010survey]. Uncertain data are not likely to represent the reality. They contain a noise that makes it deviate from its original value. This noise is mainly due to the inaccuracy and imprecision of sensors. Another source of uncertainty is the behavior of the environment, which can be unpredictable. All the computations that use uncertain data are also uncertain by propagation.

Source According to the literature, data sources are grouped into two main categories, either sensed (measured) data or computed (derived) data [DBLP:journals/comsur/PereraZCchong2007context].

Connection Context data entities are usually linked using three kinds of connections: conceptual, computational, and consistency [DBLP:conf/pervasive/HenricksenIR02; bettini2010survey]. The conceptual connection relates to (direct) relationships between entities in the real world (e.g. smart meter and concentrator). The computational connection is set up when the state of an entity can be linked to another one by a computation process (derived, predicted). Finally, the consistency connection relates entities that should have consistent values. For instance, temperature sensors belonging to the same geographical area.

Requirement

Adaptation processes aim at modifying the system state to reach an optimal one. All along this process, the system should respect the **system requirements** established ahead. Through this paper, we use the definition provided by IEEE [iso2017systems]: “(1) Statement that translates or expresses a need and its associated **constraints** and **conditions**, (2) **Condition or capability that must be met or possessed** by a

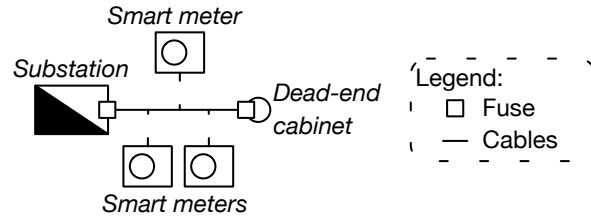


Figure 2.3: Simplified version of a smart grid

1 system [...] to satisfy an agreement, standard, specification, or other formally imposed
2 documents”.

3 Although in the literature, requirements are categorized as functional or non-functional,
4 in this paper we use a more elaborate taxonomy introduced by Glinz [DBLP:conf/re/Glinz07].
5 It classifies requirements in four categories: functional, performance, specific quality,
6 and constraint. All these categories share a common feature: they are all temporal.
7 During the life-cycle of an adaptive system, the developer can update, add or remove
8 some requirements [DBLP:conf/icse/ChengA07; pandey2010effective].

9 Action

10 In the IEEE Standards [iso2017systems], an action is defined as: “**process of**
11 **transformation** that **operates upon data** or other types of inputs to create data,
12 produce outputs, or **change the state** or condition of the subject software”.

13 Back to adaptive systems, we can define an action as a process that, given the
14 context and requirements as input, adjusts the system behavior. This modification will
15 then create new data that correspond to an output context. In the remainder of this
16 paper, we refer to output context as impacted context, or simply impact(s). Whereas
17 requirements are used to add preconditions to the actions, context information is used
18 to drive the modifications. Actions execution have a start time and a finish time. They
19 can either succeed, fail, or be canceled by an internal or external actor.

20 Use case scenario

21 In order to provide a readable and understandable example of the formalism, we
22 give a simplified version of the use case presented in Section 1.2.

Excerpt of a smart grid Figure 2.3 shows a simplified version of a smart grid with one substation, one cable, three smart meters and one dead-end cabinet. Both the substation and the cabinet have one fuse each. The meters regularly send consumption data at the same timestamp. For this example, we consider one requirement: minimizing the number of overloads. To achieve so, among the different actions, two actions are taken into account in this example: decreasing or increasing the amps limits of smart meters.

Scenario The system starts at t_0 with the actions, the requirements and all element of the context that remain fixed: the grid installation. Meters send their values at t_1 , t_2 and t_3 . Based on these data, the load on cables and substation is computed. On t_1 , an overload is detected on the cable, which breaks the requirement. At the same time point, the system decides to reduce the load of all smart meters. The impact of these actions will be measured at t_2 and t_3 , *i.e.*, the consumption will slowly reduce until the cable is no longer overloaded from t_3 .

Knowledge formalization

As discussed previously, we consider **knowledge** to be the association of **context** information, **requirements**, and **action** information, all in one global and unified model. While **context** information captures the state of the system environment and its surroundings, the system **requirements** define the constraints that the system should satisfy along the way. **Actions**, on the other hand, are meant to reach the goals of the system.

In this section, we provide a formalization of the **knowledge** used by adaptation processes based on a temporal graph. Indeed, due to the complexity and interconnectivity of system entities, graph data representation is an appropriate way to represent the **knowledge**. Augmented with a temporal dimension, temporal graphs are then able to symbolize the evolution of system entities and states over time. We benefit from the well-defined graph manipulation operations, namely temporal graph pattern matching and temporal graph relations to represent the traceability links between the **decisions** made and their **circumstances**.

Before describing this formalism, we describe the semantics used for the temporal

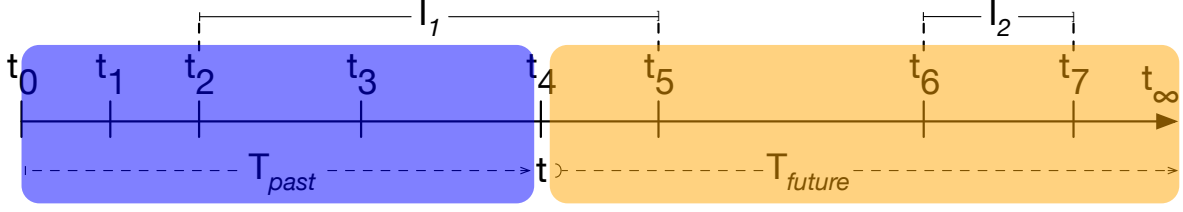


Figure 2.4: Time definition used for the knowledge formalism

axis. Then, we exemplify the knowledge formalism using the Luxembourg smart grid use case, detailed in Section 2.1.3.

Formalization of the temporal axis

The formalism described below has been made with two goals in mind. First, the definition of the time space should allow the distinction between past and future. Doing this distinction enable the differentiation between measured data and predicted (or planned data). Second, it should permit the definition of the life cycle of an element of the **knowledge**, which can be seen as a succession of states with a validity period that should not overlap each other.

Time space T is considered as an ordered discrete set of time points non-uniformly distributed. As depicted in Figure 2.4, this set can be divided into 3 different subsets $T = T_{past} \cup \{t\} \cup T_{future}$, where:

- T_{past} is the subdomain $\{t_0; t_1; \dots; t_{current-1}\}$ representing graph data history starting from t_0 , the oldest point, until the current time, t , excluded.
- $\{t\}$ is a singleton representing the current time point
- T_{future} is subdomain $\{t_{current+1}; \dots; t_{\infty}\}$ representing future time points

The three domains depend completely on the current time $\{t\}$ as these subsets slide as time passes. At any point in time, these domains never overlap: $T_{past} \cap \{t\} = \emptyset$, $T_{future} \cap \{t\} = \emptyset$, and $T_{past} \cap T_{future} = \emptyset$. The definition of these three subsets reaches the first goal.

In addition, there is a right-opened time interval $I \in T \times T$ as $[t_s, t_e)$ where $t_e - t_s > 0$. In English words, it means that the interval should represent at least one time point and should follow the time order. For any $i \in I$, $start(i)$ denotes its lower bound and

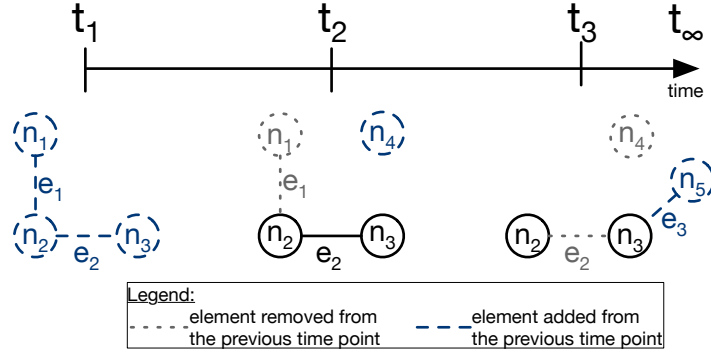


Figure 2.5: Evolution of a temporal graph over time

1 $end(i)$ its upper bound. As detailed in Section 2.2.2, these intervals are used to define
2 the validity period for each node of the graph (our second goal).

3 Figure 2.4 displays an example of a time space $T_1 = \{t_0, t_1, t_2, t_3, t_4, t_5, t_6, t_7\}$. Here,
4 the current time is $t = t_4$. According to the definition of the past subset (T_{past}) and
5 the future one (T_{future}), there is: $T_{past1} = \{t_0, t_1, t_2, t_3\}$ and $T_{future1} = \{t_5, t_6, t_7\}$. Two
6 intervals have been defined on T_1 , namely I_1 and I_2 . The first one starts at t_2 and ends
7 at t_5 and the last one is defined from t_6 to t_7 . As shown with I_1 , an interval could be
8 defined on different subsets, here it is on all of them (T_{past} , t , and T_{future}).

9 Formalism

10 **Graph definition** First, let K be an adaptive process over a system **knowledge** rep-
11 resented by a graph such as $K = (N, E)$, comprising a set of nodes N and a set of edges
12 E . Nodes represent any element of the knowledge (context, actions, *etc.*) and edges
13 represent their relationships. Nodes have a set of attribute values: $\forall n \in N, n = (id, P)$,
14 where P is the set of key-value attributes. An attribute value has a type (numerical,
15 boolean, \dots). Every relationship $e \in E$ can be considered as a couple of nodes with a
16 label $(n_s, n_t, label) \in N \times N$, where n_s is the source node and n_t is the target node.

17 **Adding the temporal dimension** In order to augment the graph with a temporal
18 dimension, the relation V^T is added. So now the knowledge K is defined as a temporal
19 graph such as $K = (N, E, V^T)$.

20 A node is considered valid either until it is removed or until one of its attributes

value changes. In the latter case, a new node with the updated value is created. Whilst, an edge is considered valid until either its source node and target node are valid, or until the edge itself is removed. Otherwise, nodes and edges are considered invalid. The temporal validity relation is defined as $V^T : N \cup E \rightarrow I$. It takes as a parameter a node or an edge ($k \in N \cup E$) and returns a time interval ($i \in I$, cf. Section 2.2.1) during which the graph element is valid.

Figure 2.5 shows an example of a temporal graph K_1 with five nodes (n_1, n_2, n_3, n_4, n_5) and three edges (e_1, e_2 , and e_3) over a lifecycle from t_1 to t_3 . In this way, K_1 equals $(\{n_1, n_2, n_3, n_4, n_5\}, \{e_1, e_2, e_3\}, V_1^T)$. Let's assume that the graph is created at t_1 . As n_1 is modified at t_2 , its validity period starts at t_1 and ends at t_2 : $V_1^T(n_1) = [t_1, t_2)$. n_2 and n_3 are not modified; their validity period thus starts at t_1 and ends at t_∞ : $V_1^T(n_2) = V_1^T(n_3) = [t_1, t_\infty)$. Regarding the edges, the first one, e_1 , is between n_1 and n_2 and the second one, e_2 from n_2 to n_3 . Both are created at t_1 . As n_1 is being modified at t_2 , its validity period goes from t_1 to t_2 : $V_1^T(e_1) = [t_1, t_2)$. e_2 is deleted at t_3 . Its validity period is thus equal to: $V_1^T(e_2) = [t_1, t_3)$.

Lifecycle of a knowledge element One node represents the state of exactly one knowledge element during a period named the validity period. The lifecycle of a knowledge element is thus modeled by a unique set of nodes. By definition, the validity periods of different nodes cannot overlap. A same time period cannot be represented by two different nodes, which could create inconsistency in the temporal graph.

To keep track of this knowledge element history, the Z^T relation is added to the graph formalism: $K = (N, E, V^T, Z^T)$. It serves to trace the updates of a given knowledge element at any point in time. This relation can also be seen as a temporal identity function which takes as parameters a given node $n \in N$ and a specific time point $t \in T$, and returns the corresponding node at that point. Formally, $Z^T : N \times T \rightarrow N$.

In order to consider this new relation in the example presented in Figure 2.5, the definition of K_1 is modified to $K_1 = (\{n_1, n_2, n_3, n_4, n_5\}, \{e_1, e_2, e_3\}, V_1^T, Z_1^T)$. In Figure 2.5, let's imagine that n_1, n_4 , and n_5 represent the same knowledge element k_e . The lifecycle of k_e is thus:

- n_1 for period $[t_1, t_2)$,

- 1 • n_4 for period $[t_2, t_3)$,
- 2 • n_5 for period $[t_3, t_\infty)$.

3 Let t'_1 be a timepoint between t_1 and t_2 . When one wants to resolve the node
4 representing the knowledge element at t'_1 , she or he gets n_1 node, no matter of the node
5 input (n_1 , n_4 , or n_5): $Z_1^T(n_4, t_1) = n_1$. On the other hand, applying the same relation
6 with another node (n_2 or n_3) returns another node. For example, if n_2 and n_3 do not
7 belong to the same knowledge element, then it will return the node given as input, for
8 example $Z_1^T(n_2, t_1) = n_2$.

9 **Knowledge elements stored in nodes** Nodes are used to store the different knowl-
10 edge elements: context, requirements and actions. The set of nodes N is thus split in
11 three subsets: $N = C \cup R \cup A$ where C is the set of nodes which store context infor-
12 mation, R a set of nodes for requirement information and A a set of nodes for action
13 information.

14 Actions define processes that indirectly impact the context: they will change the
15 behavior of the system, which will be reflected in the context information. Requirements
16 are also processes that are continuously run over the system in order to check the
17 specifications. Here, the purpose of the A and R subset is not to store these processes
18 but to list them. It can be thought as a catalogue of actions and requirements, with
19 their history.

20 Using a high-level overview, these processes can be depicted as: taking the knowl-
21 edge as input, perform tasks, and modify this knowledge as output. As detailed in the
22 next two paragraphs, action executions and requirement analysis can be formalized by
23 relations.

24 **Temporal queries for requirements** At the current state, the formalism of the
25 knowledge K does not contain any information regarding the requirement analysis. To
26 overcome this, system requirements analysis R_A are added such as $K = (N, E, V^T, Z^T,$
27 $R_A)$. R_A is a set of couples composed of patterns $P_{[t_j, t_k]}(K)$ and requirements R over
28 these patterns: $R_P = P \cup R$.

29 $P_{[t_j, t_k]}$ denotes a temporal graph pattern, where t_j and t_k are the lower and upper
30 bound of the time interval respectively. $P_{[t_j, t_k]}$ is the result of a function which takes

1 the knowledge and an interval as input: $P_{[t_j, t_k]} : K \times I$. The time interval can be either
2 fixed (absolute), *i.e.*, both bounds are precisely defined, or sliding (relative), *i.e.*, the
3 upper bound is computed from the lower bound. For example, $P_{[t_0, t_4]}$ is considered as
4 fixed and $P_{[t_0, t_0+4]}$ is considered as relative. Each element of the pattern should be valid
5 for at least one timepoint: $\forall p \in P_{[t_j, t_k]}, V^T(e) \cap [t_j, t_k) \neq \emptyset$. Patterns can be seen as
6 temporal subgraphs of K , with a time limiting constraint coming in the form of a time
7 interval.

8 **Temporal relations for actions** Like for R_A , the knowledge K needs to be aug-
9 mented with action executions A_E : $K = (N, E, V^T, Z^T, R_A, A_E)$. Actions executions
10 A_E can be regarded as a couple (A, A_F) , where A is the action that is executed and A_F
11 a set of relations or isomorphisms mapping a source temporal graph pattern $P_{[t_j, t_k]}$ to
12 a target one $P_{[t_l, t_m]}$, $A_F : K \times I \rightarrow K \times I$.

13 The left-hand side of the A_F relation depicts the temporal graph elements over which
14 an action is applied. Every relation may have a set of application conditions. They
15 describe the circumstances under which an action should take place. These application
16 conditions are either positive, should hold, or negative, should not hold. Application
17 conditions come in the form of temporal graph invariants. The side effects of these
18 actions are represented by the right-hand side.

19 Finally, we associate to A_E a temporal function E_{A_E} to determine the time interval
20 at which an action has been executed. Formally, $X : A_E \rightarrow I$.

21 **Temporal relations for decisions** Finally, the knowledge formalism needs to in-
22 clude the last, but not the least, element: decisions made by the adaptation, $K = (N,$
23 $E, V^T, Z^T, R_A, A_E, D)$ While the source of relations in D represents the state before
24 the execution of an action, the target shows its impact on the **context**. Its intent is
25 **to trace back impacts of action executions to the decisions they originated**
26 **from**.

27 A decision present in D is defined as a set of actions executed, *i.e.*, a subset of
28 A_E , combined with a set of requirement analysis, *i.e.*, a subset of R_A . Formally, $D =$
29 $\{ A_D \cup R_D \mid A_D \subseteq A_E, R_A \subseteq R_P \}$. We assume that each action should result from
30 only one decision: $\forall a \in A, \forall d1, d2 \in D \mid a \in d1 \wedge a \in d2 \rightarrow d1 = d2$.

1 The temporal function E_{A_P} is extended to decisions in order to represent the execu-
2 tion time: $E_{A_P} : (A \cup D) \rightarrow I$. For decision, the lower bound of the interval corresponds
3 to the lowest bound of the action execution intervals. Following the same principle, the
4 upper bound of the interval corresponds to the uppermost bound of the action execu-
5 tion intervals. Formally, $\forall d \in D \rightarrow E_{A_P}(d) = [l, u]$, where $l = \min_{a \in A_d} \{E_{A_P}(a)[start]\}$ and
6 $u = \max_{a \in A_d} \{E_{A_P}(a)[end]\}$.

7 **Sum up** Knowledge of an adaptive system can be formalism with a temporal graph
8 such as $K = (N, E, V^T, Z^T, R_A, A_E, D)$, wherein:

- 9 • N is a set of nodes to represent the different information (context, actions and
10 requirements)
- 11 • E is a set of edges which connects the different nodes,
- 12 • V^T is a temporal relation which defines the temporal validity of each element,
- 13 • Z^T is a relation to track the history of each knowledge elements,
- 14 • R_A is a relation that defines the different requirements processes,
- 15 • A_E is a relation that defines the different action processes,
- 16 • D is a set of action executions, which result from the same decision, and require-
17 ment analysis.

18 Decisions D can allow adaptation processes to reason over ongoing and future ex-
19 ecutions of decisions. Moreover, it allows tracing the state of the knowledge before
20 and after the decision has been or is executed, thanks to its A_D component. Plus, it
21 represents which action has been used for this. Thanks to the R_A relation, one can
22 access the requirements at the root of the decision and the state of the knowledge used
23 by this requirement.

24 In the next section, we exemplify this formalism over our case study.

25 Application on the use case

26 In this section we apply the formalism described on the use case presented in Sec-
27 tion 2.1.3.

28 Let K_{SG} be the temporal graph that represents the knowledge of this adaptive
29 system: $K_{SG} = (N_{SG}, E_{SG}, V_{SG}^T, Z_{SG}^T, R_{P_{SG}}, A_{P_{SG}}, D_{SG})$. Figure 2.6 shows the nodes

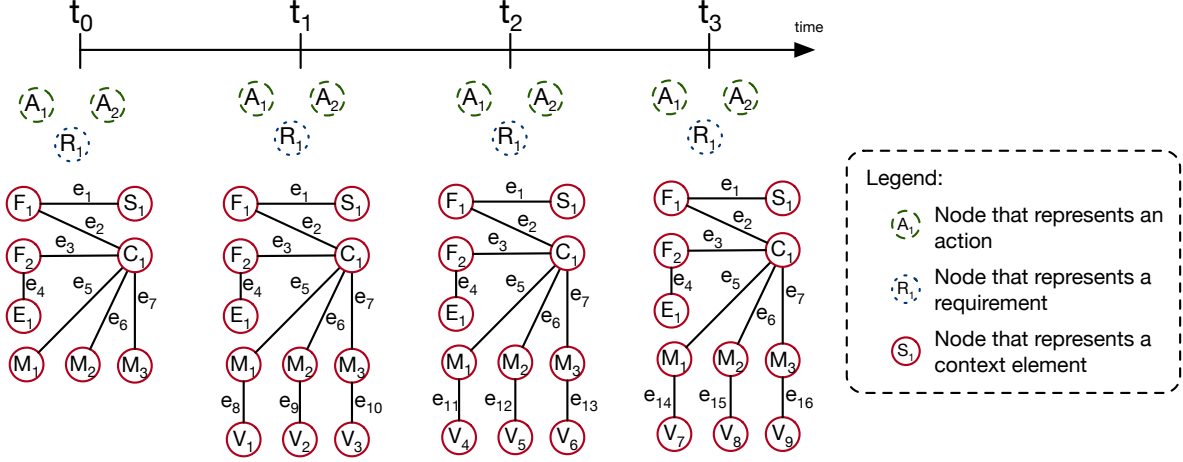


Figure 2.6: Application of the formalism with a temporal graph that represents the knowledge of the smart grid described in Section 2.1.3

- 1 and edges of this knowledge.
- 2 **Description of N_{SG}** N_{SG} is divided into three subsets: C_{SG} , R_{SG} and A_{SG} . R_{SG}
- 3 contains one node, R_1 in Figure 2.6, which represents the requirement of this example
- 4 (minimizing the number of overloads): $R_{SG} = \{R_1\}$. Two nodes, A_1 and A_2 , belong
- 5 to A_{SG} : $A_{SG} = \{A_1, A_2\}$. They represent the two actions of this example, respectively
- 6 decreasing and increasing amps limits. Regarding the context C_{SG} , there are three
- 7 nodes to represent the three smart meters (M_1 , M_2 , and M_3), one for the substation
- 8 (S_1), two for the fuses (F_1 and F_2), one for the dead-end cabinet (E_1), one for the cable
- 9 (C_1) and one node per consumption value received (V_i): $C_{SG} = \{M_1, M_2, M_3, S_1, F_1,$
- 10 $F_2, E_1, C_1\} \cup \{V_i | i \in [1..9]\}$.

11 According to the scenario, except for nodes to store consumption values, the other

12 nodes are created at t_0 and are never modified. Therefore, their validity period starts at

13 t_0 and never ends: $\forall n \in A_{SG} \cup R_{SG} \cup \{M_1, M_2, M_3, S_1, F_1, F_2, E_1, C_1\}, V_{SG}^T(n) = [t_0, t_\infty)$.

14 Considering the consumption values, all the nodes represent the history of the values

15 for the three smart meters. In other words, there are three knowledge elements: the

16 consumption measured for each meter. Let C_i notes the consumption measured by the

17 smart meter M_i . As shown in Figure ??, there is:

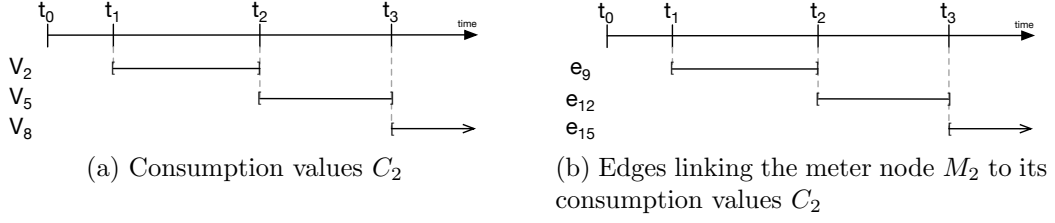


Figure 2.7: Validity periods of consumption values and their edges to the smart meter M_2

- 1 • C_1 of M_1 is represented by $\{V_1, V_4, V_7\}$,
- 2 • C_2 of M_2 is represented by $\{V_2, V_5, V_8\}$,
- 3 • C_3 of M_3 is represented by $\{V_3, V_5, V_9\}$.

4 Taking C_2 as an example, V_2 is the initial consumption value, replaced by V_5 at t_2 , itself
5 replaced by V_8 at t_3 . Applying the V_{SG}^T on these different values, results are thus:

- 6 • $V_{SG}^T(V_2) = [t_1, t_2)$,
- 7 • $V_{SG}^T(V_5) = [t_2, t_3)$,
- 8 • $V_{SG}^T(V_8) = [t_3, t_\infty)$.

9 These validity periods are shown in Figure 2.7a. As meters send the new consumption
10 values at the same time, this example can also be applied to C_1 and C_3 .

11 From these validity periods, the Z_{SG}^T can be used to navigate to the different values
12 over time. Let's continue with the same example, C_2 . In order to get the evolution of
13 the consumption value C_2 , given the initial one, one will use the Z_{SG}^T relation:

- 14 • $Z_{SG}^T(V_2, t_{s1}) = \emptyset$, where $t_0 \leq t_{s1} < t_1$
- 15 • $Z_{SG}^T(V_2, t_{s2}) = V_2$, where $t_1 \leq t_{s2} < t_2$
- 16 • $Z_{SG}^T(V_2, t_{s3}) = V_5$, where $t_2 \leq t_{s3} < t_\infty$.
- 17 • $Z_{SG}^T(V_2, t_{s4}) = V_8$, where $t_2 \leq t_{s4} < t_\infty$.

18 **Description of E_{SG}** In this example, edges are used to store the relationships between
19 the different context elements. For example, the edge between the substation S_1 and
20 the fuse F_1 allow representing the fact that the fuse is physically inside the substation.
21 Another example, edges between the cable C_1 and the meters M_1 , M_2 and M_3 represent
22 the fact that these meters are connected to the smart grid through this cable.

One may consider that relations (validity, Z^T , decisions, action executions and requirements analysis) will be stored as edges. But this decision is left to the implementation part of this formalism.

In our model, only consumption values (V_i nodes) are modified over time. Plus, since the scenario does not imply any edge modifications, only those between meters and values are modified. The edge set contains thus sixteen edges: $E_{SG} = \{e_i \mid i \in [1..16]\}$.

By definition, the unmodified edges have a validity period starting from t_0 and never ends: $\forall i \in [1..7], V_{SG}^T(e_i) = [t_0, t_\infty)$. The history of the three knowledge elements that represent consumption values do not only impact the nodes which represent the values but also the edges between those nodes and the meters ones:

- C_1 impacts edges between M_1 and V_1, V_4 , and V_7 , *i.e.*, $\{e_8, e_{11}, e_{14}\}$,
- C_2 impacts edges between M_2 and V_2, V_5 , and V_8 , *i.e.*, $\{e_9, e_{12}, e_{15}\}$,
- C_3 impacts edges between M_3 and V_3, V_6 , and V_9 , *i.e.*, $\{e_{10}, e_{13}, e_{16}\}$.

Continuing with C_2 as an example, the initial edge value is e_9 from t_1 , which is replaced by e_{12} from t_2 , itself replaced by e_{15} from t_2 . The validity relation, applied to these edges, thus returns:

- $V_{SG}^T(e_9) = [t_1, t_2) = V_{SG}^T(V_2)$,
- $V_{SG}^T(e_{12}) = [t_2, t_3) = V_{SG}^T(V_5)$,
- $V_{SG}^T(e_{15}) = [t_3, t_\infty) = V_{SG}^T(V_8)$,

These validity periods are depicted in Figure 2.7b. As they are driven by those of consumption values (V_2, V_5 , and V_8), they are equal.

As for nodes, the Z_{SG}^T relation can navigate over time through these values. For example, to get the history of the edges between the consumption value C_2 and the meter represented by M_2 , one can apply the Z_{SG}^T relation as follows:

- $Z_{SG}^T(e_9, t_{s1}) = \emptyset$, where $t_0 \leq t_{s1} < t_1$
- $Z_{SG}^T(e_9, t_{s2}) = e_9$, where $t_1 \leq t_{s2} < t_2$,
- $Z_{SG}^T(e_9, t_{s3}) = e_{12}$, where $t_2 \leq t_{s3} < t_3$,
- $Z_{SG}^T(e_9, t_{s4}) = e_{15}$, where $t_3 \leq t_{s4} < t_\infty$.

Description of D_{SG} , A_{ESG} , and R_{ASG} As described in the scenario (cf. Section 2.1.3), the requirement analysis detects that t_1 the requirement is broken. The

adaptation process will thus apply the “decreasing amps limits” action on the three meters. Following Example 2 detailed in Section 2.1.1, we consider that the action will impact the consumption values on the next two measurements: t_2 and t_3 .

In the knowledge, we thus have one decision: $D_{SG} = D_1$. This decision has been taken after one requirement analysis, R_{ASG1} , that detects no respect of the requirement R_1 . To determine if there is an overload, this analysis needs to know the topology and the consumption values. The pattern is thus defined by all nodes related to the grid network and consumption values at t_1 : $P_{1[t_1, t_1+1]} = \{S_1, F_1, F_2, C_1, E_1, M_1, M_2, M_3, V_1, V_2, V_3\}$. So we have: $R_{ASG1} = \{R_1, P_{1[t_1, t_1+1]}\}$.

The knowledge also includes the three action executions: A_{ESG1} , A_{ESG2} , and A_{ESG3} . These actions have been executed on, respectively, M_1 , M_2 , and M_3 . Following the definition, they all contain the action A_1 and similar relation which linked the circumstances to the impacts. The circumstances are the state of the knowledge at t_0 , which contain all information of the grid network and the consumption values. We denote them $P_{2[t_1, t_1+1]}$, $P_{3[t_1, t_1+1]}$, and $P_{4[t_1, t_1+1]}$, all equal $P_{1[t_1, t_1+1]}$. The impact contains all consumption values received at t_2 and t_3 . Each action impacts the consumption value of the meter that it modifies. For example, A_{ESG2} only impacts values of meter M_2 . For this action, the output pattern is thus : $P_{5[t_2, t_3]} = \{V_5, V_8\}$. In summary, A_{ESG1} , A_{ESG2} , and A_{ESG3} are defined as follows:

- for the action executed on M_1 : $A_{ESG1} = (A_1, A_{F1})$, with $A_{F1} : P_{2[t_1, t_1+1]} \rightarrow \{V_4, V_7\}$,
- for the action executed on M_2 : $A_{ESG2} = (A_1, A_{F2})$, with $A_{F2} : P_{3[t_1, t_1+1]} \rightarrow \{V_5, V_8\}$,
- for the action executed on M_3 : $A_{ESG3} = (A_1, A_{F3})$, with $A_{F3} : P_{4[t_1, t_1+1]} \rightarrow \{V_6, V_9\}$,

The decision described in the scenario is thus equal to: $D_1 = \{R_{ASG1}, A_{ESG1}, A_{ESG2}, A_{ESG3}\}$. At t_2 , this decision will still be valid. The adaptation process can thus include it in the adaptation process to reason over the ongoing actions. If at t_3 the cable remains overloaded, then one may use this element to check if the system tried to fix it, how and based on which information.

1 Actions as context information

2 In order to enhance sustainable decision-making in time evolving contexts, we
3 strongly think that actions together with their expected effects should be part of the
4 context representation of adaptive systems. In this perspective, we first describe a clus-
5 ter infrastructure management example for illustrating our approach. Then, we show
6 how an action plan can be represented as a directed acyclic graph (DAG) of actions
7 spanned over time. Then, we detail on how we enrich this DAG with the expected effect
8 of actions, and later with their measured equivalences, *i.e.*, how we link the DAG to
9 the context representation. Finally, we illustrate how this information can be used by
10 the MAPE-K loop to improve reasoning processes through four reasoning operations.

11 Example: cluster management

12 Throughout this section, we refer to a cluster management case study, the Watcher
13 project, to exemplify our approach. This framework performs adaptation processes
14 over a cloud infrastructure in order to optimize resource usage, such as CPU, memory,
15 and network. To achieve this, Watcher defines primitive actions to be performed, such
16 as container’s creation, migration, and duplication. Furthermore, Watcher comes with
17 a set of metrics, either simple or derived, for setting up optimization algorithms. For
18 instance, deciding whether to allocate or deallocate resources in the cluster. In this
19 paper, we confine to two metrics: container status and container workload. Being
20 part of the context information, these metrics values play a crucial role in planning
21 and executing adaptation processes. These values are first modified: the system will
22 update the container status and their workload. Moreover, these values can be subject
23 to changes as a result of executing the planned actions.

24 A common adaptation scenario is minimizing the amount of resources provided by
25 the cluster to answer a given service workload (W_S) that is expected a priori. The
26 framework thus has to decide whether to allocate or deallocate resources, and thereby,
27 increase or decrease the amount of the provided workload (W_P). To do so, the frame-
28 work first compares W_S to W_P , then determines the most convenient action plan. As
29 soon as an allocation or deallocation action is performed, the workload values are in-

1 creased or decreased and the container statuses are modified accordingly (e.g., *READY*,
2 *DOWN*). Neither of these actions is effective immediately and they can be subject to
3 failure. As shown in [DBLP:conf/IEEEcloud/MaoH12], a virtual machine based
4 on Linux could take from 90 seconds to more than 200 seconds to be ready.

5 Action planning

6 System adaptation is commonly performed as a set of actions. In complex scenarios,
7 these actions are often interdependent and require advanced action planning. It takes
8 into consideration not only dependency but also actions execution status. Furthermore,
9 existing simplified models, e.g., [7194653; 5984008], do not capture action execution
10 time and assume that they complete immediately after they are triggered. Capturing
11 such temporal information (e.g., (de)allocation delays) contributes to the quality of
12 context information and improves the reasoning process of the MAPE-K loop.

13 We propose to represent action plans by means of DAGs of actions spanning over
14 time, where each node corresponds to an action and an edge to a precedence relationship
15 between actions. Figure ?? shows an example of an action plan represented using an
16 actions DAG. Actions can be executed in parallel or sequential way. Each action has
17 a *start time*, an *execution duration*, and a *time until its effects are measurable* by the
18 system. These properties can be reproduced using the three temporal functions defined
19 in Section ?. The execution duration will be the difference between the *start time* and
20 the *end time* values, defined by the $start(id)$ and $end(id)$ functions. Expected effects
21 resulting from the execution of an action are also attached to actions, as highlighted
22 in Figure ?? by a dashed arrow. The *time until action effects are measurable* will
23 correspond to the start time, i.e., the result of the function $start(id)$, of the context
24 element. We differentiate between several actions statuses, namely, created, pending,
25 running, failed, succeeded, or canceled. An action is triggered when all the actions of the
26 incoming edges succeed. In case of actions failure or cancellation, the subsequent actions
27 are canceled then rescheduled for the next planning phase. We provide more details
28 about expected effects representation below. This DAG is then fed to the knowledge
29 component of the MAPE-K feedback loop to improve the reasoning process.

30 We describe in the next section how we represent this DAG and how we link it to

1 the context presentation.

2 **Linking action planning and context representation**

3 In order to enable reasoning on actions, we include two concepts in our context rep-
4 resentation model, the *measured* value and *expected* value. While *measured* represents
5 the value of data measured by the system’s sensors, *expected* represents data reflecting
6 the planned effects of actions. We assume that actions are triggered by change requests
7 made either by a human or the system itself (*e.g.*, the reasoning process). Based on
8 change requests, the reasoning process will generate the DAG of actions with effects on
9 metrics. For the sake of understandability, this concept should encapsulate the intent
10 behind performing an action.

11 An excerpt of a meta-model representing these concepts is provided in Figure ??.
12 The right-hand side of the meta-model depicts how we model change requests, thanks
13 to the *ChangeRequest* meta-class. The *name* and the *type* are used by the reasoning
14 process to infer the desired state as well as the actions to reach it. The *timeLimit*
15 attribute allows to define a time-out for each individual request.

16 The center part of the meta-model (with grey background) depicts the meta-model
17 used to create instances of the DAG of actions. To support the sequential and parallel
18 execution, we define three classes: *Action*, *Fork Action* and *Join Action*. While sequen-
19 tial actions can be modeled using the relationship *next*, parallel nodes can be modeled
20 using fork actions and join actions. The *JoinAction* is simply a node with multiple
21 incoming edges, while the *ForkAction* node with multiple outgoing edges. The actions
22 are part of the context model thanks to the reference *next* associating the *Action* class
23 to the *ContextModel* class.

24 The left-hand side deals with the sub-model that supports expected effects. A
25 value of a metric can be of any type. We represent it in the metamodel through the
26 generic type parameter *T*. Metrics with *measured values* are directly connected to the
27 context model. This corresponds to the usual semantics of context models. In the
28 Figure ??, this is shown by the *Metric* class. *ExpectedMetric* are used to represent the
29 expected values. Thanks to the *timeDephasing* and the *precision* attributes we are able
30 to handle time expressions with a high level of precision. Conceptual links between

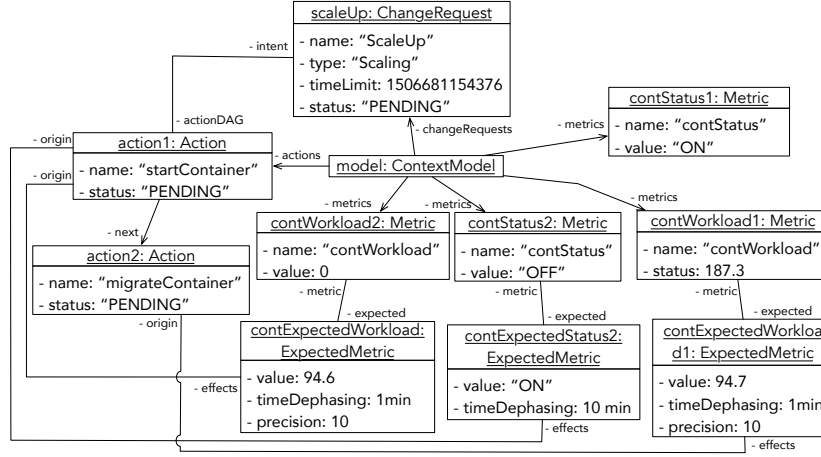


Figure 2.8: Instance model of the cluster management use cases.

1 an action and its corresponding expected metrics are established using the relationship
2 *effects* (depicted in bold red). Using this representation, we enable the analysis using
3 expected values, measured values, actions, and their effects. As we will show in the next
4 section, this representation improves significantly the analysis and planning phases of
5 traditional MAPE-K loops.

6 Coming back to our case study, Figure 2.8 depicts a model instance of our proposed
7 example. The meta-model contains four metrics: container status (*contStatus1* and
8 *contStatus2*) and container workloads (*containerWorkload1* and *containerWorkload2*).
9 All these metrics are affected by two actions: *action1*, an action to switch a container
10 on, or *action2*, an action to migrate a container to another host. As a scale up request
11 has been created, the reasoning process decides to create an action plan: first switching
12 a container on and then migrating another container.

13 Enabled reasoning

14 By associating the action model to the context model, we aim at enhancing adapta-
15 tion process with new abilities to reason. In this section, we list four kinds of reasoning
16 that is possible thanks to our approach: (i) answering change requests (ii) immediate
17 supervision, (iii) future supervision, (iv) meta supervision. We also present the neces-
18 sary code to extract information during the reasoning phase. We use a dot notation

```

model.changeRequests
  .select(ChangeRequest cr -> cr.status == "PENDING")
  .select(ChangeRequest cr -> cr.timeLimit <= NOW})
  .forEach(ChangeRequest cr -> createAction(cr))

//Rise an error for unsatisfied requests
model.changeRequests
  .select(ChangeRequest cr -> cr.status == "PENDING")
  .select(ChangeRequest cr -> cr.timeLimit > NOW)
  .forEach(ChangeRequest cr -> riseError(cr))

```

Listing 2.1: Generation of the DAG of actions from the change requested

for navigating inside our context model, *i.e.*, for accessing the properties (attributes and relationships). Moreover, we consider the elements as stream, *i.e.*, each context elements are contained in a stream of elements. For example, the metrics are contained in a stream of context elements. In addition, context elements define the same operation as the ones of the Stream Java interface⁵. The execution semantics also remains similar. When the operations are executed, they are executed at a precise time, *i.e.*, the current timestamp. When operations need to access a context element, they apply the *resolve(id,time)* functions as described in Section ?? . For readability purpose, we do not show the call to this function. Listed codes have been made using the context model presented in Figure 2.8.

Answering change requests Thanks to change requests present in the context representation, the reasoning will create a DAG of actions to reach a desired state. It thus needs to extract the change requests that have not been processed and on which the time limit has not been reached. If the time limit has been reached or overtaken, the reasoning process could throw errors. We present the code to extract the change requests from our model in Listing 2.1.

Immediate Supervision When a reasoning process plans adaptation actions, it also sets the expected effect on the metric of the context model. We refer to this operation as immediate supervision. However, the measured value could diverge from the expected one for different reasons: error in actions, context that has not changed as

⁵<https://docs.oracle.com/javase/8/docs/api/java/util/stream/Stream.html>, Last visited: 28 September 2017

expected, or wrong effect extrapolation. The reasoning process needs to detect this divergence for taking counteractions. In other words, the reasoning process needs to analyse the difference between the expected value and the measured one. We present the code to achieve the extraction of the expected and the measured value of a metric in Listing 2.2. As shown in the code, the checking should consider both the precision of the expected value and the possible temporal dephasing between the measured value and the expected effects. We check if the value is in the following range: $[effectvalue - precision; effectvalue + precision]$ and the time of the measured value is the following range: $[time - timedephasing; time + timedephasing]$.

```

10 model.metrics
11
12   .select(Metric m -> m.name == "contWorkload")
13   .forEach(Metric m -> {
14       ExpectedMetric e = m.expected;
15       if(m.value < e.value - e.precision ||
16          m.value > e.value + e.precision) {
17           if(time(m.id) < start(e.id) - e.timeDephasing ||
18              time(m.id) > start(e.id) + e.timeDephasing) {
19               createCorrectiveActions(m); }
20       }
21   })
22

```

Listing 2.2: Detection of difference between measured and expected values

Future supervision Actions are taken for adapting the system to the current context. However, they are not immediately and successfully executed and their effects are measurable with delay. As discussed in Section ??, the reasoning frequency could be inferior to this delay. It results on a reasoning process that will be executed on a similar context as the last execution. It will thus take the same action(s) whereas the context will be adapted soon. To tackle this problem, using our approach, the reasoning process can also consider the future expected metrics. By analysing the current context for detecting any miss adaptation and the future expected one for detecting any future correction, the adaptation process will avoid repetition of actions. The code below show how to extract the current container workload and the future expected one. The code remains similar to the previous one (Listing 2.2) except that we resolve the expected value at another time, here 5 min after the current time (*NOW*).


```

1 model.metrics
2   .select(Metric m -> m.name == "contWorkload")
3   .foreach(Metric m -> {
4       ExpectedMetric futureE = resolve(m.expected, NOW + 5min)
5       //similar to the code presented in Listing 1
6   })

```

Listing 2.3: Extraction of expected and measured value of a metric with a delta time

8 **Meta supervision** Supervision allows to verify that the reasoning has been suc-
9 cessfully executed, *i.e.*, that the expected effect really happened on the system. As
10 explained in the previous paragraph, the system will also diverge from this desired
11 state for different reasons. The meta supervision aims at finding extra information.
12 For example, we will detect if a desired state is not reachable, or that an action often
13 creates this divergence. The code presented in Listing 2.4 allows to detect actions that
14 did not achieve the expected effect. For this purpose, we check if an action did not
15 surpass the expected percentage *THRESHOLD* of miss expectations.

```

Map<Action,Integer> error= new HashMap();
int sum = 0;

model.timeRange(NOW - 2H, NOW)
  .metrics
  .foreach(Metric m -> {
    [...] //code of previous listing
    if(...) { // if presented in the previous listing
      Action a = m.expected.origin;
      int count = error.get(a) + 1;
      error.put(a, count);
      sum++;
    }
  });

error.forEach((key, value) -> {
  if(value / sum > THRESHOLD) {
    takeActions();
  }
});

```

Listing 2.4: Extraction of expected and measured value of a metric with a delta time

1 Modeling the knowledge

2 In order to simplify the diagnosis of adaptive systems, this thesis proposes a novel
3 **metamodel** that combines, what I call, design elements and runtime elements. Design
4 elements abstract the different elements involved in **knowledge** information to assist
5 the specification of the adaptation process. Runtime elements instead, represent the
6 data collected by the adaptation process during its execution. In order to maintain
7 the consistency between previous design elements and newly created ones, instances
8 of design elements (*e.g.*, actions) can be either added or removed. Modifying these
9 elements would consist in removing existing elements and creating new ones. Combining
10 design elements and runtime elements in the same model helps not only to acquire
11 the evolution of system but also the evolution of its structure and specification (*e.g.*
12 evolution of the requirements of the system). Design time elements are depicted in
13 gray in the Figures 2.9– 2.12. Note that, this thesis does not address how runtime
14 information is collected.

15 For the sake of modularity, the **metamodel** has been split into four packages: Knowl-
16 edge, Context, Requirement and Action. All the classes of these packages have a com-
17 mon parent class that adds the temporality dimension: *TimedElement* class. Before
18 describing the Knowledge (core) package, I detail this element. Then, I introduce in
19 more details the other three packages used by the Knowledge package: Context, Re-
20 quirement, and Action. In below sections, I use "*Package::Class*" notation to refer to
21 the provenance of a class. If the package is omitted, then the provenance package is
22 this one described by the figure or text.

23 Parent element: *TimedElement* class

24 I assume that all the classes in the different packages extend a *TimedElement* class.
25 This class contains three methods: *startTime*, *endTime*, and *modificationsTime*. The
26 first two methods allow accessing the validity interval bounds defined by the previously
27 discussed V^T relation. The last method resolves all the timestamps at which an element
28 has been modified: its history. This method is the implementation of the relation Z^T
29 described in our formalism (cf. Section ??).

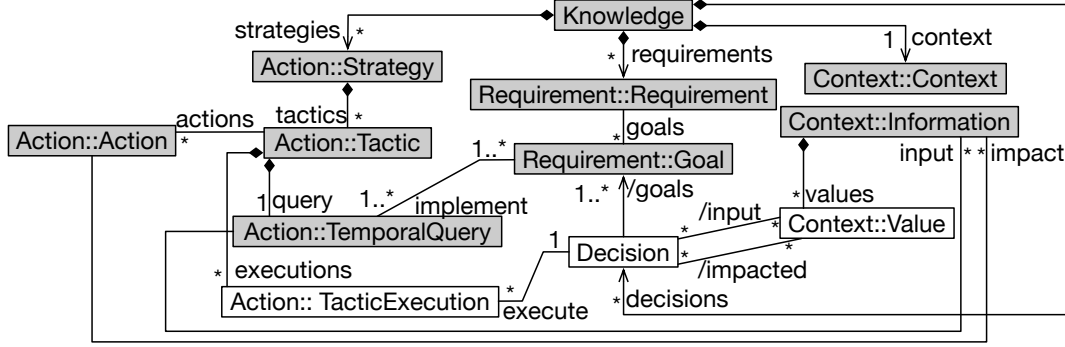


Figure 2.9: Excerpt of the knowledge metamodel

Knowledge metamodel

In order to enable interactive diagnosis of adaptive systems, traceability links between the decisions made and their circumstances should be organized in a well-structured representation. In what follows, I introduce how the **knowledge metamodel** helps to describe **decisions**, which are linked to their goals and their context (input and impact). Figure 2.9 depicts this **metamodel**.

Knowledge package is composed of a **context**, a set of **requirements**, a set of **strategies**, and a set of **decisions**. A **decision** can be seen as the output of the Analyze and Plan steps in the **Monitor, Analyze, Plan, and Execute over knowledge (MAPE-k)** loop.

Decisions comprise target *goals* and trigger the execution of one *tactic* or more. A decision has an *input* context and an *impacted* context. The context impacted by a decision (*Decision.impact*) is a derived relationship computed by aggregating the impacts of all actions belonging to a decision (see Fig. 2.12). Likewise, the *input* relationship is derived and can be computed similarly. In the smart grid example, a decision can be formulated (in plain English) as follows: since the district D is almost overloaded (*input context*), we reduce the amps limit of greedy consumers using the “*reduce amps limit*” action in order to reduce the load on the cable of the district (*impact*) and satisfy the “*no overload*” policy (*requirement*).

As all the elements inherit from the *TimedElement*, we can capture the time at which a given decision and its subsequent actions were executed, and when their impact

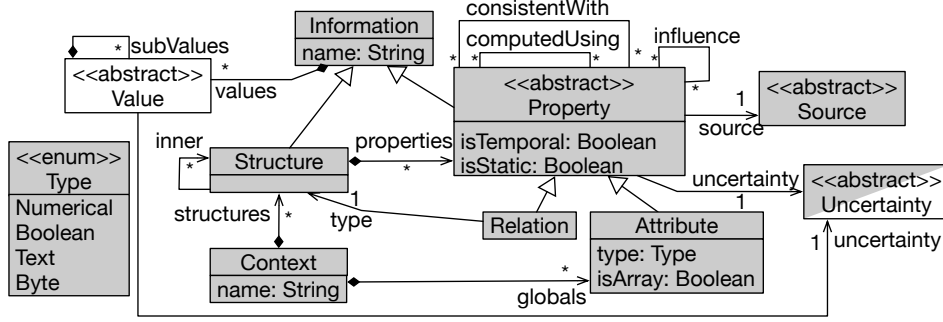


Figure 2.10: Excerpt of the context metamodel

1 materialized, *i.e.*, measured. Thanks to this metamodel representation, a developer can
 2 apprehend the possible causes behind malicious behavior by navigating from the context
 3 values to the decisions that have impacted its value (*Property.expected.impact*) and the
 4 goals it was trying to reach (*Decision.goals*). In Section **TODO: add reference** , we
 5 present an example of interactive diagnosis queries applied to the smart grid use case.

6 Context metamodel

7 Context models structure context information acquired at runtime. For example,
 8 in a smart-grid system, the context model would contain information about smart-grid
 9 users (address, names, etc.) resource consumption, etc.

10 An excerpt of the context model is depicted in Figure 2.10. I propose to rep-
 11 resent the context as a set of structures (*Context.structures*) and global attributes
 12 (*Context.globals*). A structure can be viewed as a C-structure with a set of properties
 13 (*Property*): attributes (*Attribute*) or relationships (*Relation*). A structure may contain
 14 other nested structures (*Structure.inner*). Structures and properties have values. They
 15 correspond to the nodes described in the formalization section (*cf.* Section ??). The
 16 connection feature described in Section 2.1.2 is represented thanks to three recursive
 17 relationships on the Property class: *consistentWith*, *computedUsing* and *influence*. Ad-
 18 ditionally, each property has a source (*Source*) and an uncertainty (*Uncertainty*). It
 19 is up to the stakeholder to extend data with the appropriate source: measured, com-
 20 puted, provided by a user, or by another system (*e.g.*, weather information coming

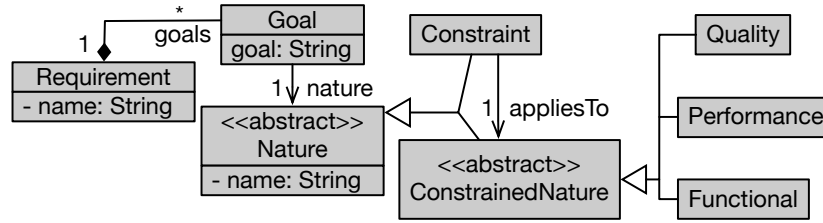


Figure 2.11: Requirement metamodel

1 from a public API). Similarly, the uncertainty class can be extended to represent the
2 different kinds of uncertainties. Finally, a property can be either historic or static.

3 Requirement metamodel

4 As different solutions to model system requirements exist (*e.g.*, KAOS [dardenne1993goal],
5 i* [yu2011modelling] or Tropos [DBLP:journals/aamas/BrescianiPGGM04]),
6 in this metamodel, we abstract their shared concepts. The requirement model, de-
7 picted in Figure 2.11, represents the *requirement* as a set of *goals*. Each goal has a
8 *nature* and a textual specification. The nature of the goals adheres to the four cat-
9 egories of requirements presented in Section 2.1.2. One may use one of the existing
10 requirements modeling languages (*e.g.*, RELAX) to define the semantics of the require-
11 ments. Since the requirement model is composed solely of design elements, we may rely
12 on static analysis techniques to infer the requirement model from existing specifications.
13 The work of Egyed [egyed01] is one solution among others. This work is out of the
14 scope of the paper and envisaged for future work.

15 In the guidance example, the requirement model may contain a **balanced resource**
16 **distribution** requirement. It can be split into different goals: (i) *no overload*, (ii) *no*
17 *production lack*, (iii) *no production loss*.

18 Action metamodel

19 Similar to the requirements metamodel, the actions metamodel also abstracts main
20 concepts shared among existing solutions to describe adaptation processes and how they
21 are linked to the context. Figure 2.12 depicts an excerpt of the action metamodel. I de-
22 fine a strategy as a set of tactics (*Strategy*). A tactic contains a set of actions (*Action*).

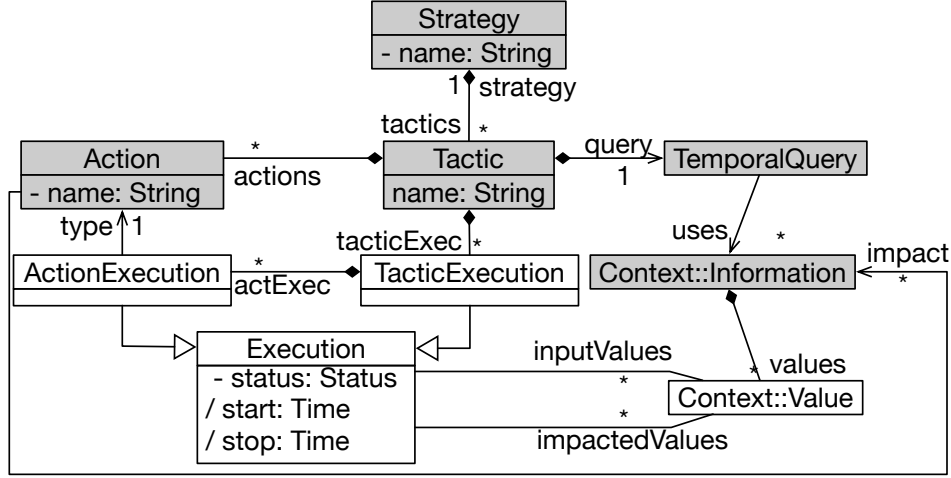


Figure 2.12: Excerpt of the action metamodel

1 A tactic is executed under a precondition represented as a temporal query (*Temporal-*
2 *Query*) and uses different data from the context as input. In future work, we will inves-
3 tigate the use of preconditions to schedule the executions order of the actions, similarly
4 to existing formalisms such as Stitch [DBLP:journals/jss/ChengG12]. The query
5 can be as complex as needed and can navigate through the whole knowledge model.
6 Actions have impacts on certain properties, represented by the *impacted* reference.

7 The different executions are represented thanks to the *Execution* class. Each ex-
8 ecution has a status to track its progress and links to the impacted context val-
9 ues(*Execution.impactValues*). Similarly, input values are represented thanks to the
10 *Execution.inputValues* relationship. An execution has *start* and *end* time. Not to con-
11 fuse with the *startTime* and *endTime* of the validity relation \mathcal{V}^T . Whilst the former
12 corresponds to the time range in which a value is valid, the *start* and *stop* time in the
13 class execution correspond to the time range in which an action or a tactic was being
14 executed. The start and stop attributes correspond to the relation \mathcal{X} (see Section ??).
15 These values can be derived based on the validity relation. They correspond to the
16 time range in which the status of the execution is "RUNNING". Formally, for every
17 execution node e , $\mathcal{X}(e) = (\mathcal{V}(e) \mid e.status = "RUNNING")$.

18 Similarly to requirement models, it is possible to automatically infer design elements

1 of action models by statically analyzing actions specification. Since acquiring informa-
2 tion about tactics and actions executions happens at runtime, one way to achieve this is
3 by intercepting calls to actions executions and updating the appropriate action model
4 elements accordingly. This is out of the scope of this paper and planned for future
5 work.

6 Validation

7 Validation of action consideration

8 In this paper, we argue that by adding action information directly within context
9 representations the reasoning phase (*i.e.*, the *Analyse and Plan* of the MAPE-k loop)
10 is improved. We validate the proposal using our case study example, the cloud man-
11 agement scenario. We demonstrate the scalability of our approach could by applying
12 it to large-scale adaptive systems.

13 To implement our approach and build this experimentation we used the GreyCat
14 framework⁶. It is a models@run.time framework with time as a built-in concept[DBLP:phd/basesearch
15 The code for this experimentation is open source and publicly accessible⁷.

16 Validation

17 In our experimentation, we simulate a cluster that executes a unique service on
18 several containers. The adaptation process allocates or deallocates resources in/from
19 the cluster infrastructure by relying on the measured workload of resource usage. The
20 objective function is to minimize the number of used resources while reducing the
21 difference between W_S and W_P . The targeted state is $W_S = W_P$.

22 We use the fastStorage⁸ dataset[DBLP:conf/ccgrid/ShenBI15] to simulate the
23 service workload W_S . To reproduce a service workload, we sum up the CPU usage in
24 the dataset. Then, the workload of the containers W_C is computed by dividing W_S by
25 the number of available containers N_C . A maximum of 8,600⁹ is set up to represent

⁶<http://greycat.ai/>, Last visited: 21 September 2017

⁷<https://bitbucket.org/ludovicpapers/sac-eval>

⁸<http://gwa.ewi.tudelft.nl/datasets/gwa-t-12-bitbrains>, Last visited: 22 September 2017

⁹Average of allocated CPUs in MHz in the dataset

1 the allocated CPU of containers. Due to this maximum constraint, the sum of the
 2 container workloads ($W_P = \sum_{i=0}^{N_C} \leq W_S$), also referred to as the provided workload
 3 of the service W_P , could be inferior to the workload of the service. In addition, the
 4 maximum value of the provided workload W_{Pmax} is equal to the sum of the maximum
 5 of the container: $W_{Pmax} = N_C * 8600$.

6 We compare the output results of similar adaptation processes but different sets of
 7 rules. The first set uses a traditional rule-based decision-making process [**aamodt1994case**;
 8 **anicic2010rule**] while the second one replicates the rules and takes into consideration
 9 the expected effects of actions.

10 We identify two rules:

- 11 • **Rule 1** if $W_P < W_S$ then start as many containers as needed
- 12 • **Rule 2** if $W_{Pmax} > W_S$ then stop as many containers as needed

13 The equivalent rules in a temporal-aware context are depicted below:

- 14 • **Rule 1** if $W_P(now) < W_S(now)$ and $W_E(now + 3min) < W_S(now)$ then start
 15 $(W_S(now) - W_E(now + 3min))/8600$ containers
- 16 • **Rule 2** if $W_{Pmax}(now) > W_S(now)$ and $W_{Pmax}(now + 3min) > W_S(now)$ then
 17 stop $(W_{Pmax}(now + 3min) - W_S(now))/8600$ containers

18 $W_E(t)$ represents the expected workload of a service at a time t, which is also the
 19 sum of the expected workload ($W_{CE}(t)$) of all the containers at the time t. We formalize
 20 it as: $W_E(t) = \sum_{i=0}^{N_C} W_{CE_i}(t)$.

21 The adaptation process is executed every 1 minute, for 100 times, in order to simu-
 22 late the reactivity of the clusters. We show that rules that ignore the expected effects
 23 of actions perform poorly. We compute the number of start actions and the number
 24 of stop actions against the two sets of rules. Using the first set of rules, we reach a
 25 total of 433 start actions and 515 stop actions whereas using our approach 168 start
 26 actions and 235 stop actions. Thus, the second set of rules reduces the number of start
 27 and stop actions by a factor of ~ 2.6 and ~ 2.2 respectively. The difference between the
 28 provided workload and the service workload remains similar in the two usages.

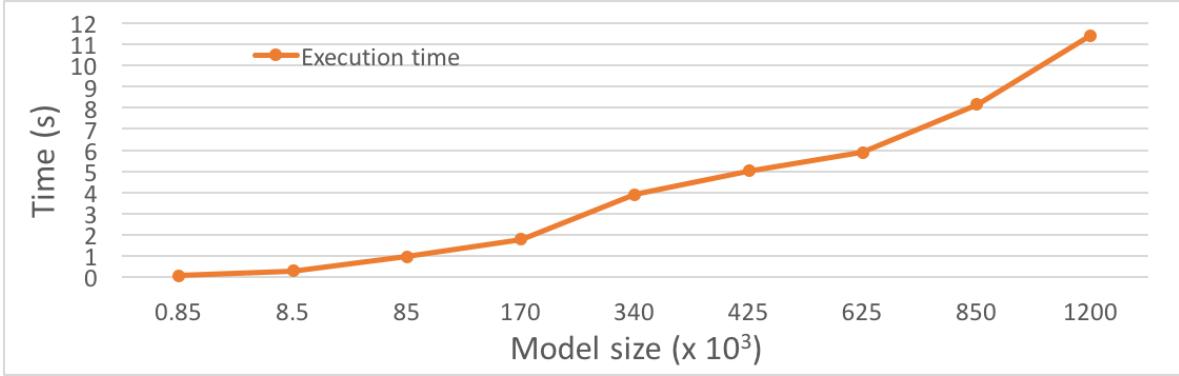


Figure 2.13: Latency to detect faults within temporal contexts given different sizes of cloud clusters

1 Evaluation of reasoning latency

In this section, we evaluate the efficiency of our context representation for enhancing a reasoner implementing the previously defined rules (2nd set of rules). All the experimentations have been executed in a MacBook Pro with an Intel Core i7 processor (2.8GHz, 4 cores, 16GB main memory (RAM), macOS Sierra version 10.12.4) using Oracle JDK 1.8.0_65. Our proposed temporal context representation introduces inescapably an overhead compared to a simple vector of metrics, as it is commonly used in cloud management implementations. Using an experimental evaluation, we answer this question by varying the size of the context model in order to simulate various sizes of clusters. Then, we evaluate during this experiment the latency of our cloud model-based reasoner to trigger corrective actions. The results are presented in Figure 2.13. The depicted results have been obtained by computing the average execution time over ten (10) runs.

Experimentally, we obtained a very good latency (*less than 200ms*, which is comparable to a single Remote Procedure Call [mulligan2009comparison]) for a cloud cluster with fewer than 1000 containers. For larger clusters, the latency remains acceptable (*less than 1s*) until reaching 85,000 elements. Finally, we evaluated a very large-scale cluster (*1,200,000 containers*), where we obtained a latency of less than 12s. In the later case, we obtained a throughput of 100,000 rules/second. The performance

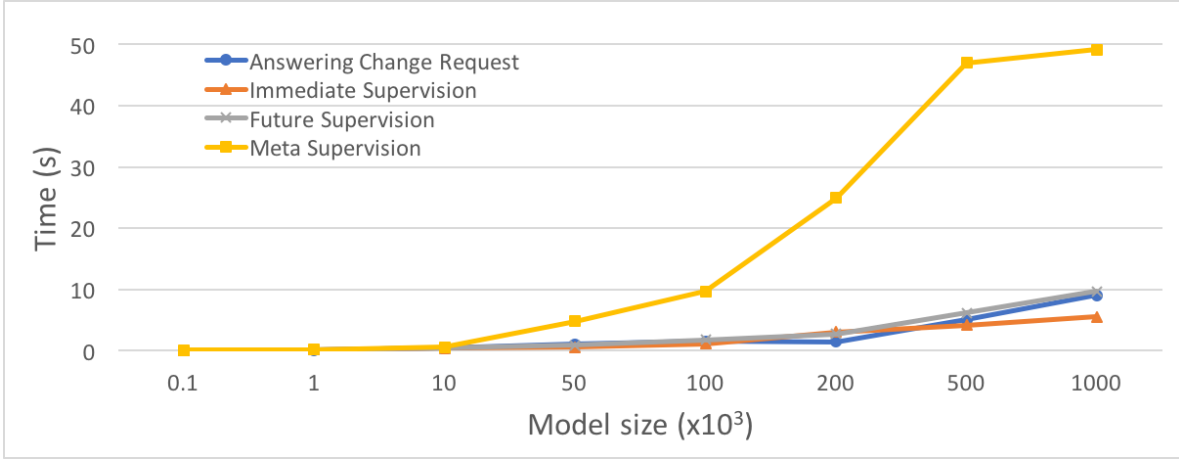


Figure 2.14: Execution time of the new reasoning abilities

of our approach can be considered as very good as cloud clusters rarely go beyond few thousands of elements [DBLP:conf/ccgrid/ShenBI15].

Evaluation of our reasoning element

In the last experimentation, we evaluate the four reasoning capacities presented in Section 2.3. Figure 2.14 depicts the average results of this experimentation of ten (10) executions. The first three reasoning operations, answering change requests, immediate supervision, and future supervision, give very good execution time. These operations can be evaluated these three operations over 1 million of elements in less than 10 seconds. As shown in Table 2.1, it results in a throughput of more than 100,000 operations per second. However, our results are less good for the meta-supervision. We can execute more than 20,000 of operations per second. This is due to the computational complexity of this supervision. Contrary to the previous reasoning that are executed at a precise time, the meta supervision is executed over a time range. But, the meta supervision operations should not be used with the same frequency as the three other ones. It could be used with less frequency as it analyses the data over bigger time range.

Table 2.1: Reasoning operation throughput

Reasoning operation	Throughput (op/s)
Answering change requests	110,702.74
Immediate Supervision	157,748.61
Future supervision	102,409.70
Meta supervision	20,345.88

Validation for the temporal graph

This section briefly presents the implementation of our approach on top of an existing framework to store temporal graphs. We validate our approach by first describing how one can implement graph manipulation operations involved in diagnosis algorithms. Later, we evaluate the performance of our implementation..

Application on the smart grid example

To validate and evaluate our approach, we implemented a prototype publicly available online¹⁰. This implementation leverages the GreyCat framework¹¹, more precisely the modeling plugin, which allows designing a metamodel using a textual syntax. Based on this specification, GreyCat generates a Java and a JavaScript API to create and manipulate models that conform to the predefined metamodel. The GreyCat framework handles time as a built-in concept. Additionally, it has a native support of a lazy loading mechanism and an advanced garbage collection. This is achieved by dynamically loading and unloading model elements from the main memory when necessary.

In what follows, we explain how a stakeholder, Morgan, can apply our approach to a smart grid system in order to, first, abstract adaptive system concepts, then, structure runtime data, and finally, query the model for diagnosis purpose. The corresponding object model is depicted in Figure 2.15. Due to space limitation, we only present an excerpt of the knowledge model. An elaborate version is accessible in the tool repository.

Abstracting the adaptive system At design time (t_d), either manually or using an automatic process, Morgan abstracts the different tactics and actions available in

¹⁰<https://github.com/lmouline/LDAS>

¹¹<https://github.com/datathings/greycat>

the adaptation process. Among the different tactics that Morgan would like to model is “*reduce amps limit*”. It is composed of three actions: sending a request to the smart meter (*askReduce*), checking if the new limit corresponds to the desired one (*checkNewLimit*), and notifying the user by e-mail (*notifyUser*). Morgan assumes that the *askReduce* action impacts consumption data (*csmp*). This tactic is triggered upon a query (*tempQ*) that uses meter (*mt*), consumption (*csmp*) and customer (*cust*) data. The query implements the “*no overload*” goal: the system shall never have a cable overload. Figure 2.15 depicts a flattened version of the temporal model representing these elements. The tag at upper-left corner of every object illustrates the creation timestamp. All the elements created at this stage are tagged with t_d .

Adding runtime information The adaptation process checks if the current system state fulfills the requirements by analyzing the context. To perform this, it executes the different temporal queries, including *tempQ*. For some reasons, the *tempQ* reveals that the current context does not respect the “*no overload*” goal. To adapt the smart grid system, the adaptation process decides to start the execution of the previously described tactic (*exec1*) at t_s . As a result, a decision element is added to the model along with a relationship to the unsatisfied goal. In addition, this decision entails the planning of a tactic execution, manifested in the creation of the element *exec1* and its subsequent actions (*notifyU*, *checkLmt*, and *askRed*). At t_s , all the actions execution have an IDLE status and an expected start time. All the elements created at this stage are tagged with the t_s timestamp in Figure 2.15.

At t_{s+1} , the planned tactic starts being executed by running the action *askReduce*. The status of this action turns from *IDLE* to *RUNNING*. Later, at t_{s+2} , the execution of *askReduce* finishes with a *SUCCEED* status and triggers the execution of the actions *notifyUser* and *checkNewLimit* in parallel. The status of *askReduce* changes to *SUCCEED* while the status of *notifyUser* and *checkNewLimit* turns to *RUNNING*. The first action successfully ends at t_{s+3} while the second ends at t_{s+4} . As all actions terminates with a *SUCCEED* status at t_{s+4} , accordingly, the final status of the tactic is set *SUCCEED* and the *stop* attribute value is set to t_e .

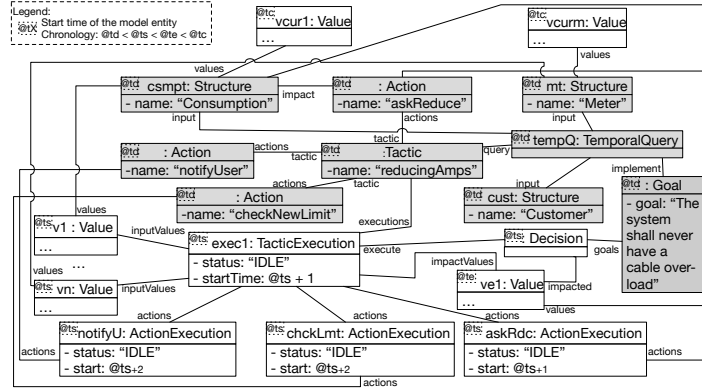


Figure 2.15: Excerpt of the knowledge object model related to our smart grid example

1 **Interactive diagnosis query** After receiving incident reports concerning regular
2 power cuts, and based on the aforementioned knowledge model, Morgan would be able
3 to query the system's states and investigate why such incidents have occurred. As
4 described in Section ??, she/he will interactively diagnose the system by interrogating
5 the context, the decisions made, and their circumstances.

6 The first function, depicted in Listing 2.5, allows to navigate from the currently
7 measured values (*vcurl*) to the decision(s) made. The for-loop and the if-condition are
8 responsible for resolving the measured data for the past two days. Past elements are
9 accessed using the *resolve* function that implements the \mathcal{Z}^T relation (cf. Section ??).
10 After extracting the decisions leading to power cuts, Morgan carries on with the diag-
11 nosis by accessing the circumstances of this decision. The code to perform this task
12 is depicted in Listing 2.5, the second function (*getCircumstances*). Note that the rela-
13 tionship *Decision.input* is the aggregation of *Decision.execute.inputValues*.

```

14 // extracting the decisions
15 Decision [] impactedBy(Value v) {
16     Decision [] respD
17     for( Time t : v.modificationTimes() ):
18         if ( t >= v.startTime() - 2 day )
19             Value resV = resolve(v, t)
20             respD.addAll(from(resV).navigate(Value.impactd))
21     return respD
22 }
23
24 // extracting the circumstances of the made decisions

```

```

1 Tuple<Value [], Goal[]> getCircumstance(Decision d) {
2   Value [] resValues = from(d).navigate(Decision.input)
3   Goal [] resGoals = from(d).navigate(Decision.goals)
4   return Tuple<>(resValues, resGoals)
5 }

```

Listing 2.5: Get the goals used by the adaptation process from executed actions

7 Performance evaluation

8 GreyCat stores temporal graph elements in several key/value maps. Thus, the com-
9 plexity of accessing a graph element is linear and depends on the size of the graph.
10 Note that in our experimentation we evaluate only the execution performance of di-
11 agnosis algorithms. For more information on I/O performance in GreyCat, please
12 refer to the original work by Hartmann *et al.*, [DBLP:conf/seke/0001FJRT17;
13 DBLP:phd/basearch/Hartmann16].

```

14 MATCH (input)-[*4]->(output)
15 WHERE input.id IN [randomly generated set]
16 RETURN output
17 LIMIT 0

```

Listing 2.6: Traversal used during the experimentations

20 We consider a diagnosis algorithm to be a graph navigation from a set of nodes
21 (input) to another set of nodes (output). Unlike typical graph algorithms, diagnosis
22 algorithms are simple graph traversals and do not involve complex computations at
23 the node level. Hence, we believe that three parameters can impact their performance
24 (memory and/or CPU): the global size of the graph, the size of the input, and the
25 number of traversed elements. In our evaluation, we altered these parameters and report
26 on the behavior of the main memory and the execution time. The code of our evaluation
27 is publicly available online¹². All experiments reporting on memory consumption were
28 executed 20 times after one warm-up round. Whilst, execution time experiments were
29 run 100 times after 20 warm-up rounds. The presented results correspond to the mean
30 of all the iterations. We randomly generate graph with sizes (N) ranging from 1 000
31 to 2 000 000. At every execution iteration, we follow these steps: (1) in a graph with

¹²<https://bitbucket.org/ludovicpapers/icac18-eval>

size N , we randomly select a set of I input nodes, (2) then traverse M nodes in the graph, (3) and we collect the first O nodes that are at four hops from the input element. Listing 2.6 describes the behavior of the traversal using Cypher, a well-known graph traversal language.

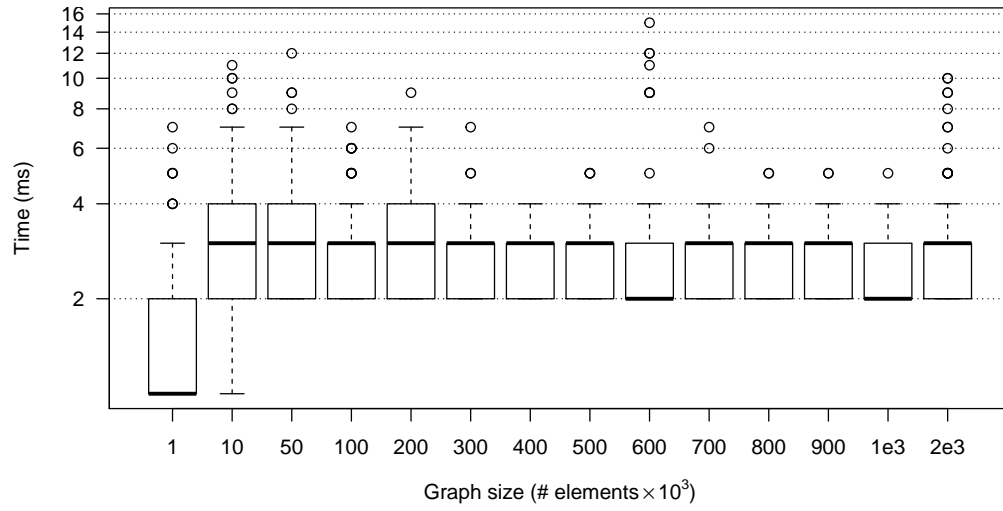
We executed our experimentation on a MacBook Pro with an Intel Core i7 processor (2.6 GHz, 4 cores, 16GB main memory (RAM), macOS High Sierra version 10.13.2). We used the Oracle JDK version 1.8.0.65.

How performance is influenced by the graph size N ? This experimentation aims at showing the impact of the graph size (N) on memory and execution time while performing common diagnosis routines. We fix the size of I to 10. To assure that the behavior of our traversals is the same, we use a seed value to select the starting input elements. We stop the algorithm when we reach 10 elements. Results are depicted in Figure 2.16.

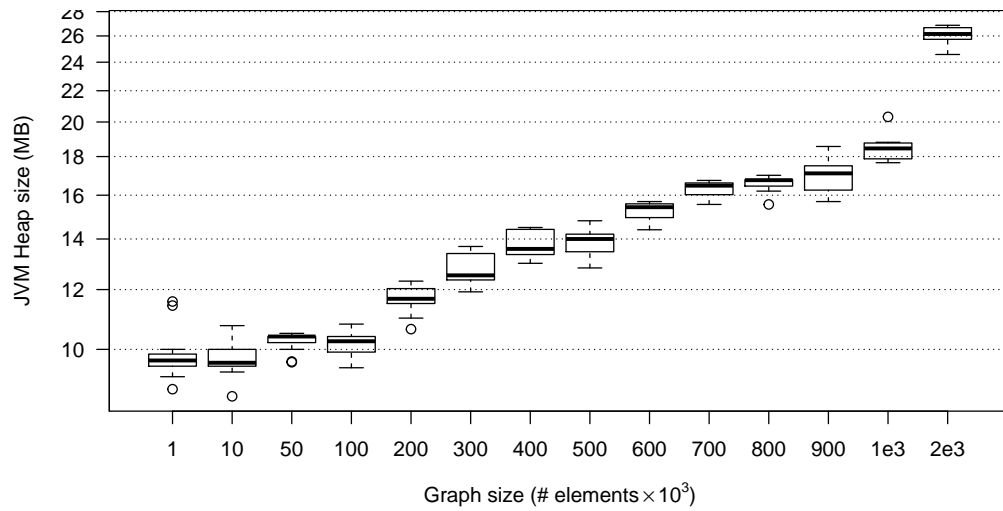
As we can notice, the graph size does not have a significant impact on the execution time of diagnosis algorithms. For graphs with up to 2,000,000 elements, execution time remains between 2 ms and four 4 ms. We can also notice that the memory consumption insignificantly increases. Thanks to the implementation of a lazy loading and a garbage collection strategy by GreyCat, the graph size does not influence memory or execution time performance. The increase in memory consumption can be due to the internal indexes or stores that grow with the graph size.

How performance is influenced by the input size (I)? The second experiment aims to show the impact of the input size (I) on the execution of diagnosis algorithms. We fix the size of N to 500 000 and we variate I from 1 000 nodes to 100 000, *i.e.*, from 0.2% to 20% of the graph size. The results are depicted in Figure 2.17 (straight lines).

Unlike to the previous experiment, we notice that the input size (I) impacts the performance, both in terms of memory consumption and execution time. This is because our framework keeps in memory all the traversed elements, namely the input elements. The increase in memory consumption follows a linear trend with regards to N . As it can be noticed, it reaches 2GB for $I=100\,000$. The execution time also shows a similar curve, while the query response time takes around than around 60ms to run for $I=1\,000$,



(a) Execution time evolution



(b) Memory evolution

Figure 2.16: Experimentation results when the knowledge based size increases

1 it takes a bit more than 4 seconds to finish for $I=100\,000$. Nonetheless, these results
2 remain very acceptable for diagnosis purposes.

3 **How performance is influenced by the number of traversed elements (M)?**

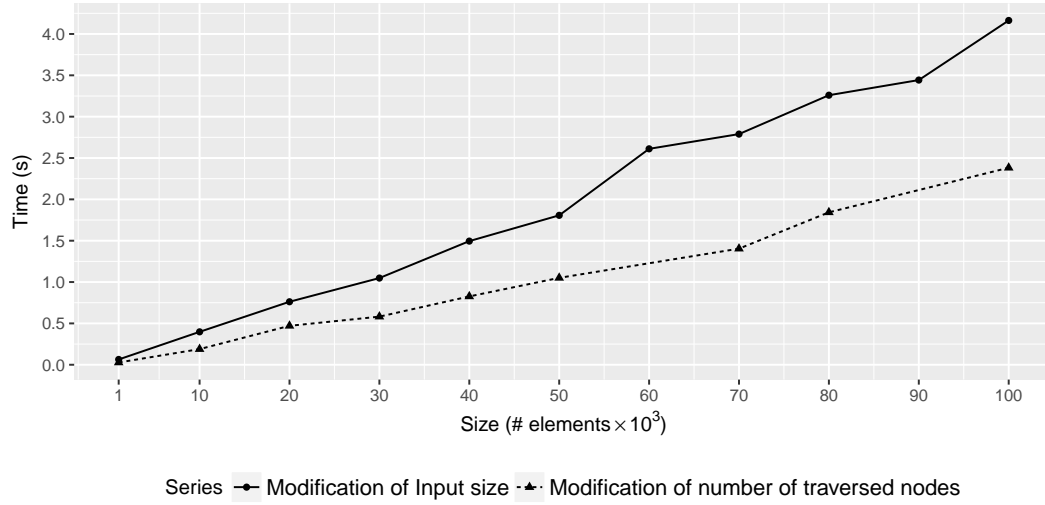
4 For the last experiment, we aim to highlight the impact of the number of traversed
5 elements (M). For this, we fix I and O to 1, and randomly generate a graph with
6 sizes ranging from 1 000 to 100 000. Our algorithm navigates the whole model ($M=N$).
7 We depict the results in Figure 2.17 (dashed curve). As we can notice, the memory
8 consumption increases in a quasi-linear way. The memory footprint to traverse $M =$
9 100 000 elements is around 0.9GB. The progress of the execution time curve behaves
10 similarly, in a quasi-linear way. Finally, the execution time of a full traversal over the
11 biggest graph takes less than 2.5 seconds.

12 **Discussion**

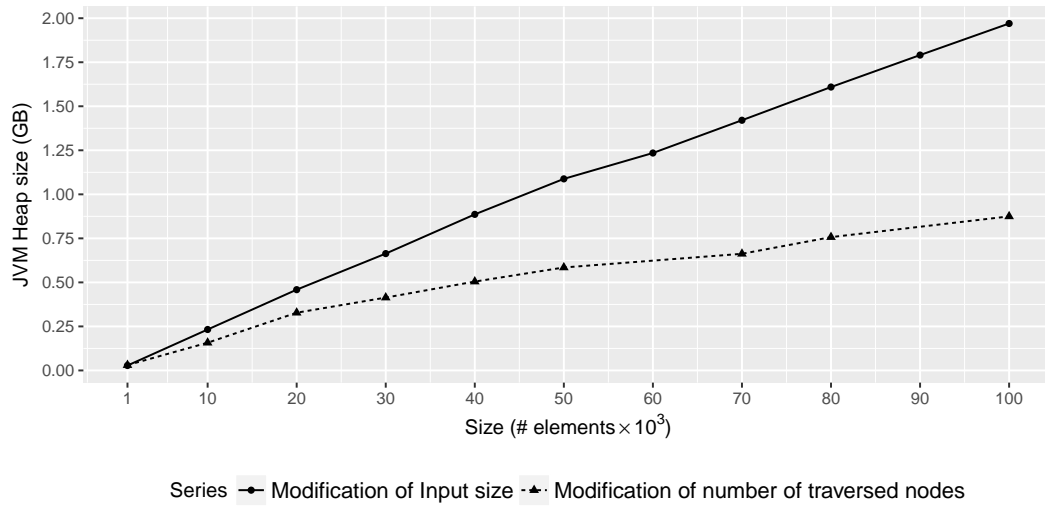
13 By linking context, actions, and requirements using decisions, data extraction for
14 explanation or fault localization can be achieved by performing common temporal graph
15 traversal operations. In the detailed example, we show how a stakeholder could use our
16 approach to define the different elements required by such systems, to structure runtime
17 data, finally, to diagnose the behavior of adaptation processes.

18 Our implementation allows to dynamically load and release nodes during the execu-
19 tion of a graph traversal. Using this feature, only the needed elements are kept in the
20 main memory. Hence, we can perform interactive diagnosis routines on large graphs
21 with an acceptable memory footprint. However, the performance of our solution, in
22 terms of memory and execution time, is restricted by the number of traversed elements
23 and the number of input elements. Indeed, as shown in our experimentation, both the
24 execution time and the memory consumption grow linearly.

25 As described in [DBLP:conf/smartgridcomm/0001FKTPTR14], the smart
26 grid in Luxembourg is composed of 1 central system, 3 data concentrators and 227 me-
27 ters. The network is thus composed of 231 elements. Each meter sends the consumption
28 value every 15 min, being 908 every hours. Plus, there is from 0 to 273 topology mod-
29 ifications in the network. In total, the system generates from 908 to 1,181 new values
30 every hour. If we consider that we have one model element per smart grid entity and



(a) Evolution of the execution time



(b) Evolution of the memory consumption

Figure 2.17: Results of experiments when the number of traversed or input elements increases

1 one model element per new value, 100,000 model elements correspond thus from $((100,$
2 $000 - 231) * 1H) / 1,181 = 84,5H$ ($\sim 3,5$ days) to $((100,000 - 231) * 1H) / 908 = 109,9H$
3 ($\sim 4,6$ days) of data. In other word, our approach can efficiently interrogate up to ~ 5
4 days history data in 2.4s.

5 **Threat to validity**

6 **Conclusion**

¹ Abbreviations

² **MAPE-k** Monitor, Analyze, Plan, and Execute over knowledge. ³⁴, *Glossary*: **MAPE-**
³ **k**

1 Glossary

2 **action** “Process that, given the **context** and **requirements** as input, adjusts the system
3 behavior”, IEEE Standards [iso2017systems]. 15

4 **circumstance** State of the **knowledge** when a **decision** has been taken. 15

5 **context** In this document, I use the definition provided by Anind K. Dey [DBLP:journals/puc/Dey01
6 “Context is any information that can be used to characterize the situation of an entity.
7 An entity is a person, place, or object that is considered relevant to the interaction
8 between a user and [the system], including the user and [the system] themselves”. 15,
9 20, 34

10 **decision** A set of **actions** taken after comparing the state of the **knowledge** with the
11 **requirement**. 15, 34

12 **knowledge** The knowledge of an adaptive system gathers information about the **con-**
13 **text**, **actions** and **requirements**. 15–17, 33, 34

14 **MAPE-k** A theoretical model of the adaptation process proposed by Kephart and
15 Chess [DBLP:journals/computer/KephartC03]. It divides the process in four
16 stages: monitoring, analysing, planning and executing. These four stages share a **knowl-**
17 **edge**. 34, *Abbreviation:* MAPE-k

18 **metamodel** Through this thesis, I use the definition of Seidewitz: “A metamodel
19 is a specification model for a class of [system under study] where each [system un-
20 der study] in the class is it-self a valid model expressed in a certain modeling lan-
21 guage.” [DBLP:journals/software/Seidewitz03] . 33, 34

1 **requirement** “(1) Statement that translates or expresses a need and its associated
2 constraints and conditions, (2) Condition or capability that must be met or possessed
3 by a system [...] to satisfy an agreement, standard, specification, or other formally
4 imposed documents”, IEEE Standards [iso2017systems]. 15, 34