Online Detection of Operator Errors in Cloud Computing Using Anti-patterns

Arthur Vetter

Online Detection Of Operator Errors In Cloud Computing Using Anti-Patterns

Arthur Vetter

Horus software GmbH, Ettlingen, Germany
arthur.vetter@horus.biz

Abstract. *IT services are subject of several maintenance operations like upgrades, reconfigurations or redeployments. Monitoring those changes is crucial to detect operator errors, which are a main source of service failures. Another challenge, which exacerbates operator errors is the increasing frequency of changes, e.g. because of continuous deployments like often performed in cloud computing. In this paper, we propose a monitoring approach to detect operator errors online in real-time by using complex event processing and anti-patterns. The basis of the monitoring approach is a novel business process modelling method, combining TOSCA and Petri nets. This model is used to derive pattern instances, which are input for a complex event processing engine in order to analyze them against the generated events of the monitored applications.*

Keywords: Complex Event Processing, Anti-Pattern, TOSCA, IT Service Management, Anomaly Detection

1 Introduction

Operator errors have been one of the major reasons for IT service failures [1]-[6] and will probably continue to be regarding current trends like continuous delivery, DevOps and infrastructure-as-code [7], which are especially very common in cloud computing. In recent years, several studies and methods were developed to detect errors in very complex IT systems [8]. Those traditional methods are suited for detecting errors during ”normal” operations, but not during change operations like reconfigurations or rolling upgrades, when one node after the other is upgraded [9]. The reason for the amount of operator errors is their human nature, because change operations are either performed or initiated by human operators.

This paper presents current research results of a novel monitoring approach for those change operations. The monitoring approach is based on a process model, combining TOSCA (Topology and Orchestration Specification for Cloud Applications) and high-level Petri nets [8], which explicitly models the maintenance operations of the IT service applications. This process model is used to derive pattern instances from it. Those pattern instances are checked through a complex event processing engine against state events and transaction events. State events describe the state of the application, whereas transaction events describe each single operation performed on the application. Therefore, the logs of the applications are filtered for meaningful transaction events and are sent to the complex event processing engine, allowing the detection of operator errors almost in real-time. The complex event processing engine compares the pattern instances with the generated events through anti-patterns and creates an error message, when an anti-pattern instance was detected. Figure 1 gives an overview of the general monitoring approach.

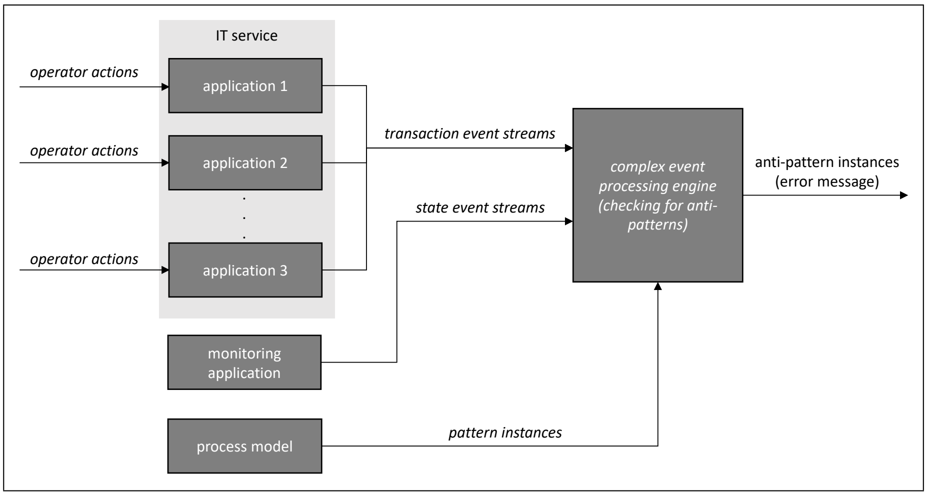


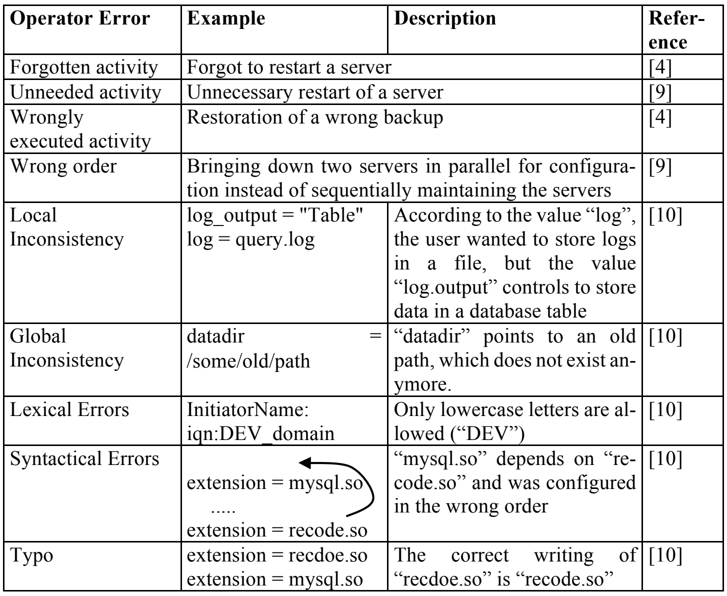
Fig. 1. General monitoring approach

The remainder of this paper is organized as follows: The next section gives a short overview of typical operator errors. Section three describes the fundamentals of TOSCA and XML nets, which are used to model the actual maintenance. Section four describes the concept of patterns and anti-patterns. Section five presents the proof of concept implementation. In the next section first experimental results are presented and discussed in the following section. Afterwards related work is presented. Section nine concludes the paper.

2 Operator Errors

Oppenheimer et al. [5] and many other authors like [4], [5], [11], [12] classify operator errors in process errors and configuration errors. Process errors can be further differentiated in following errors: forgotten activity, an unneeded activity was executed, a wrong activity was executed or actual correct activities were executed in the wrong order. Configuration errors can be separated in formatting errors and configuration value errors [13]. Formatting errors can be further separated in lexical errors, syntactical errors and typos. Configuration value errors can be further classified in local value inconsistencies and global environment inconsistencies. A monitoring approach to detect operator errors should be able to detect all those process and configuration error types. Table 1 gives an example for every type of operator error and a reference to a study with further information and examples.

Table 1. Operator error examples



Operator Error	**Example**	**Description**	**Reference**
Forgotten activity	Forgot to restart a server		[4]
Unneeded activity	Unnecessary restart of a server		[9]
Wrongly executed activity	Restoration of a wrong backup		[4]
Wrong order	Bringing down two servers in parallel for configuration instead of sequentially maintaining the servers		[9]

Local Inconsistency	log_output = "Table"
	log = query.log
		According to the value "log", the user wanted to store logs in a file, but the value "log.output" controls to store data in a database table	[10]

Global Inconsistency	datadir = /some/old/path	"datadir" points to an old path, which does not exist anymore.	[10]
Lexical Errors	InitiatorName: iqn:DEV_domain	Only lowercase letters are allowed ("DEV")	[10]

Syntactical Errors	extension = mysql.so
	.....
	extension = recode.so
		"mysql.so" depends on "recode.so" and was configured in the wrong order	[10]

Typo	extension = recdoe.so
	extension = mysql.so
		The correct writing of "recdoe.so" is "recode.so"	[10]

3 Fundamentals

The process model is a combination of TOSCA and XML nets and was introduced in a former paper [8]. In this chapter, we describe the fundamentals of TOSCA and XML nets shortly and then describe how maintenance operations can be modelled with TOSCA and XML nets.

3.1 TOSCA

TOSCA is a standard, released by OASIS [14] to support the portability of cloud applications between different cloud providers and the automation of cloud application provisioning. Therefore, TOSCA provides a modelling language to describe cloud applications as Service Templates. A Service Template consists of a Topology Template and of optional Plans (see figure 2).

A Topology Template describes the structure of a cloud application as a directed graph and consists of Node Templates and Relationship Templates. A

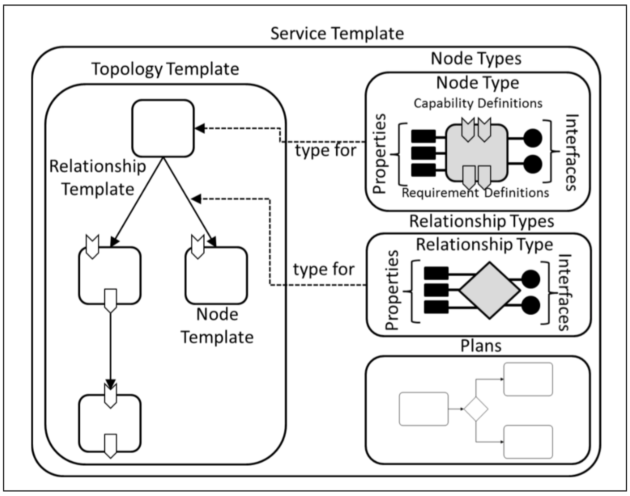


Fig. 2. TOSCA Service Template

Node Template represents a component of the cloud application, e.g. an application server and is described by a Node Type. A Node Type defines

- properties of the component (*Properties Definition*),
- available operations to manipulate the component (*Interfaces*),
- requirements of the component (*Requirement Definitions*),
- possible lifecycle states of the component (*Instance States*) and
- capabilities it offers to satisfy other components' requirements (*Capability Definitions*).

Plans are models to orchestrate the management Operations, which are offered by the cloud application components and can be written in BPMN, BPEL or other languages.
As TOSCA Service Templates are written as XML documents, we decided to use the notation of XML nets for the creation of Plans, which we name ”maintenance plan” in the rest of the paper. Using XML nets has the advantage that no additional notation elements have to be defined like it is the case e.g. for BPMN [25]. Apart of that, XML nets allow to describe detailed manipulations of XML documents, which are used to model configuration operations in maintenance plans.

3.2 XML nets

XML nets [15] are a high-level variant of Petri nets, in which places represent containers for XML documents. The XML documents must conform to the XML Schema, which is assigned to a specific place. Edges are labeled with Filter Schemas, which are used to read or manipulate XML documents. Transitions can be inscribed by a logical expression, whose variables are contained in the adjacent edges. A transition in an XML net is enabled and can be fired for a given marking, when the following three conditions hold. First, every place in the pre-set of the transition holds at least one valid XML document, which conforms to the Filter Schema inscribing the edge to the transition. Second, every place in the post-set of a transition must contain one valid XML document, if the XML document has to be modified. If an XML document has to be created from scratch the place must not already contain this XML document. Third, for the given instantiation of the variables, the transition inscription has to be evaluated to true in order to enable the transition. If an enabled transition is fired, XML documents in the pre-set places are (partially) deleted or read for the given instantiation of variables, and new XML documents are created or existing XML documents are modified in the post-set places of the transition.

3.3 Modelling maintenance plans

This section describes the modelling of maintenance plans with TOSCA and XML nets, which allows to model applications and the orchestration of applications' management operations in one integrated model. Such a model can then be used to derive pattern instances. Therefore, we extend our former approach, introduced in [8]. The following adjustments are made to the general definition of TOSCA Node Templates:

- A Node Template represents exactly one instance of an application, that means the attributes *minIstances, maxInstances:=1.*
- Node Templates are extended with the complex element *InstanceState*, which stores the current state of the corresponding application.

The notation of XML nets is adjusted as follows:

- Places are containers for *Service Templates.* Every place is assigned to the general TOSCA XML schema and additionally to a single *Node Type*, which restricts the allowed filter schemas for corresponding *Node Templates.*
- Transitions represent operations, defined in *Interfaces* of the adjacent *Node Types.*
- Filter Schemas can either be used to select *Node Templates* or to modify *Properties*, or *Instance States* of a *Node Template.* Deleting whole *Node Templates* is in contrast to general XML nets not allowed. *Node Templates* can only change their status, e.g. to undeploy, but they cannot be deleted. The reason is, that for error detection purposes, even an undeployed Node has to be monitored to be sure it was really undeployed and e.g. has not been deployed by accident afterwards again. Deleting parts of a *Node Template* are allowed, e.g. deleting a property.
- Transitions hold the attributes *start* and *end*, which define when the operation has to be executed earliest and latest.

We define a maintenance plan as a tuple MP = $<\mathrm{P,T,A},\mathit{\Psi},\mathrm{I_P,I_N,I_A,I_T,M_0}>$, where

(i) <P,T,A>is a Petri net with a set of places P, a set T of transitions, and a set A of edges connecting places and transitions (the definition and description of petri nets is excluded in this paper, but can be found, e.g., in [11]).

(ii) Ψ = <D,FT,PR >is a structure consisting of a finite and non-empty individual set D, a set of term and formula functions *FT* defined on *D*, and a set of predicates *PR* defined on *D*.

(iii) $\mathrm{I_P}$ is the function that assigns the TOSCA XML Schema to each place.

(iv) $\mathrm{I_N}$ is the function that assigns additionally a Node Type to each place.

(v) $\mathrm{I_A}$ is the function that assigns a Filter Schema to each edge. The Filter Schema must conform to the XML Schema and Node Type of the adjacent place.

(v) $\mathrm{I_T}$ is the function that assigns a predicate logical expression as inscription to each transition. The inscription is built on a given structure Ψ and a set of variables. Only variables, which are contained in the Filter schemas of adjacent arcs, are allowed. The inscription must evaluate to true in order to enable the transition.

(vi) Each transition represents a value of the element *operation*, which is defined in the complex element *Interfaces* of the Node Type in the postset of the transition.

(vii) $\mathrm{M_0}$ is the initial marking. Markings are TOSCA Service Templates.

(viii) Each transition holds the attributes *start* and *end*.

Figure 3 shows an example of a maintenance plan to configure the database connection of the application *MyApplication* (Filter Schemas are written informally for readability reasons). It is assumed that the database and application are part of the Service Template *MyService*. *MyApplication* is hosted on *MyAppServer* and requires additionally the database *TestDatabase* (NT1). It is assumed, that when the change is performed, *MyApplication* is started. In the first place, which is linked to a Node Type *Application*, *MyApplication* is one possible representation. The first Filter Schema *FS1* selects *MyApplication*. Before *MyApplication* can be configured it has to be stopped, which is represented in the first transition. The condition in order to stop the application is, that *MyApplication* has to be started. Stopping is one possible operation, which is given by the Node Type *Application*. If at the beginning of executing the change, *MyApplication* is already stopped instead of started, it is a hint, that an incident or something unexpected happened, so the change execution should be interrupted. When *MyApplication* is stopped, the database connection can be set. Therefore, the Node Template *TestDatabase* is selected and the database connection is built up on the properties of *TestDatabase* and inserted in *MyApplication* through the Filter Schema *FS5*. Afterwards *MyApplication* can be started again, but only if *TestDatabase* is running (inscription assigned to transition *Start*).

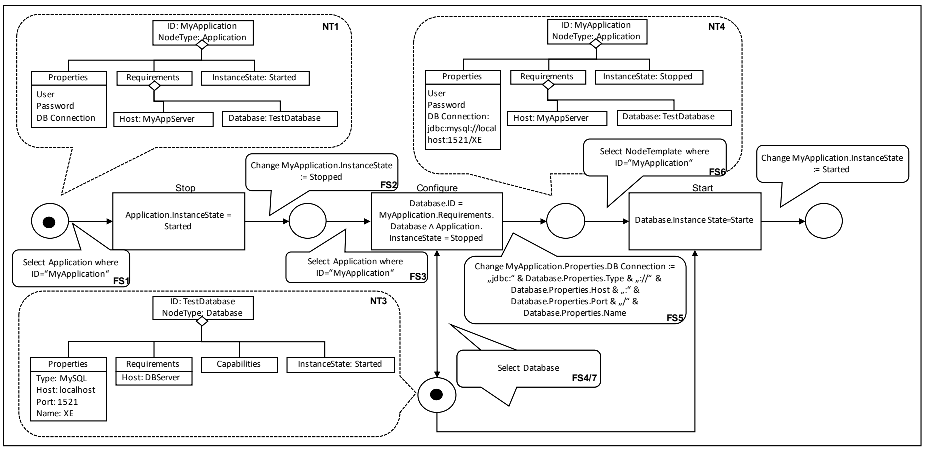

ID: MyApplication
NodeType: Application
NT1
Properties
User
Password
DB Connection
Requirements
InstanceState: Started
Host: MyAppServer
Database: TestDatabase
ID: MyApplication
NodeType: Application
NT4
Properties
User
Password
DB Connection:
jdbc:mysql://local
host:1521/XE
Requirements
InstanceState: Stopped
Host: MyAppServer
Database: TestDatabase
Change MyApplication.InstanceState := Stopped
FS2
Select NodeTemplate where ID="MyApplication"
FS6
Change MyApplication.InstanceState := Started
Stop
Application.InstanceState = Started
Configure
Database.ID = MyApplication.Requirements. Database ∧ Application. InstanceState = Stopped
Start
Database.Instance State=Starte
Select Application where ID="MyApplication"
FS1
Select Application where ID="MyApplication"
FS3
Change MyApplication.Properties.DB Connection := „jdbc:" & Database.Properties.Type & „://" & Database.Properties.Host & „:" & Database.Properties.Port & „/" & Database.Properties.Name
FS5
ID: TestDatabase
NodeType: Database
NT3
Properties
Type: MySQL
Host: localhost
Port: 1521
Name: XE
Requirements
Host: DBServer
Capabilities
InstanceState: Started
Select Database
FS4/7

Fig. 3. Example of a TOSCA based XML net

4 Pattern and Anti-Pattern for operator error detection

In computer science the term pattern is popular since the publication of the book about design patterns from Gamma et al. [12]. In this book, Gamma et al. describe patterns as solutions for recurring problems in a specific context. Aalst et al. [13] used the concept of patterns for business process modelling and described several patterns for the control flow perspective. Since then, many patterns were described for different perspectives of business process modelling, like for the data perspective [14], [15]. Riehle and Züllighoven define a pattern more general as an abstraction of a recurring concrete form in a specific context [16]. A form is a finite number of distinguishable elements and their relationships [16]. A context restricts the possible usage of a form, because the form has to fit into this specific context. Based on the definition of Riehle and Züllighoven we define a pattern and anti-pattern as following:

DEFINITION 4.1: (PATTERN). A pattern is an abstraction of a welcomed, recurring, concrete form in a specific context.

DEFINITION 4.2: (ANTI-PATTERN). An anti-pattern is as an abstraction of an unwelcomed, concrete form in a specific context.

In our work, we use patterns to describe the planned to be control flow, application configurations and application states for the scheduled maintenance. So, patterns are used during the design phase. Anti-patterns are used to check during the actual execution of the maintenance (run-time), if a form of events exists, which does not fit to the planned forms. In the following we restrict and formalize the context of the used patterns and anti-patterns as well as the form of these patterns and anti-patterns.

4.1 Context

As described in chapter three, the monitoring approach is based on the comparison between produced events of monitored applications and pattern instances of the TOSCA management plan. Those parameters build the context of the patterns. We separate two kinds of events in our context: *state events* and *transaction events*. Definitions 4.3 and 4.4 formalize *state events* and *transaction events* in this paper.

DEFINITON 4.3: (STATE EVENT). A state event is a tuple se=(timestamp, app, state), where:

- *timestamp* is the timestamp of the event creation.
- *app* is the *Node Template id* of the monitored application.
- *state* is the actual state of the application. Only values are allowed, which are defined in the Node Type of the application by the element *Instance States*.

The set of all state events is defined as *SES*.

DEFINITON 4.4: (TRANSACTION EVENT). A transaction event is a tuple *te=(timestamp, st, app, op, prop, value)*, where:

- *timestamp* is the timestamp of the event creation.
- *st* is the *Service Template id*, which identifies the service the application belongs to.
- *app* is the Node Template id of the monitored application.
- *op* describes the operation, which was conducted on the application. The value of *op* must correspond to one of the values, which are defined in the element *operation* of the *Node Type* of the application.
- *prop* describes the property, which was changed when the operation was executed. If no property was changed during the operation *prop* is null.
- *value* is the value of the property, which was changed. If *prop* is null, *value* also has to be null.

The set of all transaction events is defined as *TES*. State events and transaction events represent the actual events during a maintenance. The corresponding "to-be" events are conditions and activities, which can be derived from a TOSCA management plan. A condition represents a possible transition inscription, whereas activities represent firing sequences.

DEFINITON 4.5: (CONDITION). A condition is a tuple *(app, op, prop, zapp, state)*, where:

- *app* is the id of the Node Template, on which the operation is performed.
- *op* is the operation, which is performed on the Node Template and is restricted in the Node Type of the Node Template.
- *prop* is the property of the Node Template, which is changed during the operation.

- *zapp* is the id of the Node Template, which has to be in a specific state in order to perform the operation.
- *state* describes in which state *zapp* has to be.

Let *SM* be the set of all maintenance plans. The set of all conditions of a maintenance plan is defined as $\mathrm{SC_i}$, $\mathrm{i} \in \mathrm{SM}$. The set of all transition inscriptions of a maintenance plan is defined as $\mathrm{STI_i}$ $\mathrm{i} \in \mathrm{SM}$. The function F: $\mathrm{SC_i} \rightarrow \mathrm{STI_i}$ assigns a transition to each condition.

DEFINITON 4.6: (ACTIVITY). An activity is a tuple *a=(st, app, op, prop, value, start, end)*, where:

- *st* is the *Service Template id*, which identifies the service template in the TOSCA management plan.
- *app* is the id of the Node Template, on which the operation is performed.
- *op* is the operation, which is performed on the Node Template and is restricted in the Node Type of the Node Template.
- *prop* is the property of the Node Template, which is changed during the operation.
- *value* is the value of the property, which was changes. If prop is null, value also has to be null.
- *start* describes when the activity has to start earliest.
- *end* describes when the activity has to end latest.

Be *SM* the set of all maintenance plans. The set of all activities of a maintenance plan is defined as $\mathrm{SA_i}$, $\mathrm{i} \in \mathrm{SM}$. The set of all transitions of a maintenance plan is defined as $\mathrm{ST_i}$, $\in \mathrm{SM}$. The function F: $\mathrm{SA_i} \rightarrow \mathrm{ST_i}$ assigns a transition to each activity.

Additionally, for some anti-patterns we need the history of transaction events and the latest state of an application called the state event history.

DEFINITON 4.7: (TRANSACTION EVENT HISTORY). A *transaction event history* is a selection on the set of transaction events, which are in the time scope of the scheduled maintenance:
$\mathrm{TEH} := \sigma_{\mathrm{timestamp \geq maintenance\_start} \ \wedge \ \mathrm{timestamp \leq maintenance\_end}} \mathrm{TES}$

DEFINITON 4.8: (STATE EVENT HISTORY). The *state event history* SEH stores the latest state for each application in *SES*.

Furthermore, we define three functions, time, countTE and countA.

DEFINITON 4.9: (TIME). time is a function, which returns the current timestamp.

DEFINITON 4.10: (COUNTTE). *countTE(te,TEH)* is a function, which counts the number of occurrences of the transaction event *te* in the transaction event history.

DEFINITON 4.11: (COUNTA). *countA(a, S)* is a function, which counts the number of occurrences of an activity *a* in a set S.

After the description and definition of the context, the patterns and anti-patterns are described.

4.2 Pattern and Anti-Pattern

All in all, we define ten patterns/anti-patterns in order to detect operation errors. These are NEXT, IMMEDIATELY NEXT, PRECEDENCE, IMMEDIATELY PRECEDENCE, OCCURRENCE, ALTERNATIVE OCCURRENCE, ABSENCE, ALTERNATIVE ABSENCE, VALUE and STATE-CONDITION. The first eight patterns are highly influenced by the specification pattern of Dwyer at al. [17] and are used to detect process errors. Whereas the VALUE anti-pattern is used to detect configuration errors. The STATE-CONDITION anti-pattern is used to check, if a resource is in the planned state in order to perform a task on it. To describe the patterns and anti-patterns following template is used:

- **Name**: The name of the pattern must be unique and should describe the purpose of the pattern.
- **Description**: Here the form of the pattern is described, which should occur in the maintenance.
- **Instances**: Here it is described, how instances of the pattern can be derived from the maintenance plan.
- **Example**: Here, examples of pattern instances are given.
- **Anti-pattern**: A description of the corresponding anti-pattern and which type of operator errors can be detected with the anti-pattern. Additionally, we formalize the conditions, which have to be violated in order to detect an operator error.
- **Similar pattern**: Here, similar patterns are referenced and differences are named.

Pattern NEXT
Description: This pattern describes pairs of activities, defining which activity has to occur after another (with possible activities inbetween). The pattern is used for controlling AND-joins, AND-splits and concurrent sequences in a maintenance plan.
Instances: To get all instances of this pattern for a TOSCA management plan i we create a relation P1 := $\mathrm{AM_i}$ x $\mathrm{AM_i}$ x $\mathrm{AM_i}$ with the tuples ($\mathrm{a_{cur}}$, $\mathrm{a_{nex}}$, $\mathrm{a_{far}}$) where,

- the corresponding transitions $\mathrm{t_{cur}}$ and $\mathrm{t_{nex}}$ of the activities $\mathrm{a_{cur}}$ and $\mathrm{a_{nex}}$ are connected through the same place,

- $\mathrm{t}_{\mathrm{cur}}$, $\mathrm{t}_{\mathrm{nex}}$ und $\mathrm{t}_{\mathrm{far}}$ have to occur in the same path,
- $\mathrm{t}_{\mathrm{far}}$ always has to occur after $\mathrm{t}_{\mathrm{cur}}$,
- $\mathrm{t}_{\mathrm{far}}$ and $\mathrm{t}_{\mathrm{cur}}$ may not be connected through the same place.

Example: In figure 4 instances of the pattern NEXT are (a1, a2, a4), (a1, a2, a6), (a1, a3, a5), (a1, a3, a7), (a2, a4, a6) and (a3, a5, a7).

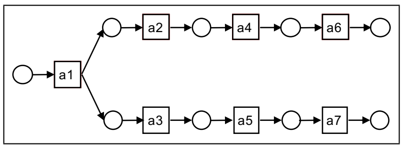


Fig. 4. Example pattern NEXT

Anti-pattern: The anti-pattern allows to detect operator errors of the type ”wrong order”. Besides, it is possible to detect operator errors of the type syntactical error, if a configuration parameter was changed in the wrong order. An error message is created, when a transaction event $\mathrm{te}_{\mathrm{cur}}$ in the event stream occurs and none of the next events $\mathrm{te}_{\mathrm{nex}}$ conforms to the next activity $\mathrm{a}_{\mathrm{nex}}$. However, one of the next events conforms to an activity $\mathrm{a}_{\mathrm{far}}$:

$$\begin{gathered}
\pi_{\mathrm{app,op,prop}} \mathrm{te}_{\mathrm{cur}} \in \pi_{\mathrm{a_{cur}.app,a_{cur}.op,a_{cur}.prop}} \mathrm{P1_i} \succ \pi_{\mathrm{app,op,prop}}\ \mathrm{te}_{\mathrm{nex}} \in \\
\pi_{\mathrm{app,op,prop}}\ (\pi_{\mathrm{a_{far}}}(\sigma_{\mathrm{a_{cur}.app=te_{cur}.app}\ \wedge\ \mathrm{a_{cur}.op=te_{cur}.op}\ \wedge\ \mathrm{a_{cur}.prop=_{cur}.prop}} \mathrm{P1_i})) \wedge \\
\pi_{\mathrm{app,op,prop}}(\pi_{\mathrm{a_{nex}}}\ (\\
\sigma_{\mathrm{a_{cur}.app=te_{cur}.app}\ \wedge\ \mathrm{cur.op=te_{cur}.op}\ \wedge\ \mathrm{a_{cur}.prop=te_{cur}.prop}\ \wedge} \\
{}_{\mathrm{a_{far}.app=te_{nex}.app}\ \wedge\ \mathrm{a_{far}.op=te_{nex}.op}\ \wedge\ \mathrm{a_{far}.prop=te_{nex}.prop}} \mathrm{P1_i})) \notin \\
\pi_{\mathrm{app,op,prop}}(\sigma_{\mathrm{te_{cur}.timestamp}\ >\ \mathrm{timestamp}\ \wedge\ \mathrm{te_{nex}.timestamp}\ <\ \mathrm{timestamp}} \mathrm{TEH})
\end{gathered}$$

Similar pattern: The pattern IMMEDIATELY NEXT allows also to detect operator errors of the type ”wrong order”. However, the pattern IMMEDIATELY NEXT would create wrong error messages for concurrent sequences and can only be used for non-concurrent activities.

Pattern IMMEDIATELY NEXT

Description: This pattern describes pairs of activities, defininig which activity has to be executed right after another activity (without any other activities occuring inbetween). The pattern is used for controlling XOR-joins, XOR-splits and non-concurrent sequences in a maintenance plan.

Instances: To get all instances of this pattern for a maintenance plan i we create a relation $\mathrm{P2_i} := \mathrm{AM_i}\ \mathrm{x}\ \mathrm{AM_i}$ with the tuples ($\mathrm{a}_{\mathrm{cur}}$, $\mathrm{a}_{\mathrm{nex}}$) where,

- the corresponding transitions of the activities $\mathrm{a}_{\mathrm{cur}}$ and $\mathrm{ta}_{\mathrm{nex}}$ are connected through the same place, and
- the corresponing transitions cannot be executed concurrent to other transitions.

Example: In figure 5 instances of the pattern IMMEDIATELY NEXT are (a1, a2), (a1, a3), (a2, a4), (a3, a5), (a4, a6) and (a5, a7).

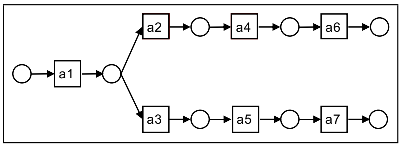


Fig. 5. Example pattern IMMEDIATELY NEXT

Anti-pattern: The anti-pattern allows to detect operator errors of the type "wrong order" for non-concurrent operations. An error message is created when a transaction event $\mathrm{te}_{\mathrm{cur}}$ occurs in the event stream and the next following transaction event $\mathrm{te}_{\mathrm{cur+1}}$ of the same service template does not correspond to the expected activity:

$$\pi_{\mathrm{app,op,prop}}\mathrm{te}_{\mathrm{cur}} \in \pi_{\mathrm{a_{cur}.app,a_{cur}.op,a_{cur}.prop}}\mathrm{P2_i} \succ \pi_{\mathrm{app,op,prop}}(\sigma_{\mathrm{st=te_{cur}.st}}\mathrm{te}_{\mathrm{cur+1}}) \notin \pi_{\mathrm{app,op,prop}}(\pi_{\mathrm{a_{nex}}}(\sigma_{\mathrm{a_{cur}.app=te_{cur}.app}\ \wedge\ \mathrm{a_{cur}.op=te_{cur}.op}\ \wedge\ \mathrm{a_{cur}.prop=te_{cur}.prop}}\mathrm{P2_i}))$$

Similar pattern: The pattern IMMEDIATELY NEXT is similar to the pattern NEXT. The difference is, that in the IMMEDIATELY NEXT pattern in contrast to the pattern NEXT no activites of the same service template are allowed between a pair of activities.

Pattern PRECEDENCE

Description: This pattern describes pairs of activities where one activity has to occur before another one. Like the pattern NEXT it is allowed that other activities occur inbetween the activitites of such a pair of activities. The pattern is used for controlling AND-joins, AND-splits and concurrent sequences in a maintenance plan.

Instances: To get all instances of the pattern a relation $\mathrm{P3_i} := \mathrm{AM_i} \times \mathrm{AM_i}$ with the tuples ($\mathrm{a_{cur}}$, $\mathrm{a_{pre}}$) is created where,

- the corresponding transitions of the activities $\mathrm{a_{cur}}$ and $\mathrm{a_{pre}}$ are connected through the same place, and
- the corresponing transitions can be executed concurrent to other transitions.

Example: In figure 4 instances of the pattern PRECEDENCE are (a2, a1), (a3, a1), (a4, a2), (a5, a3), (a6, a4) and (a7, a5).

Anti-pattern: With the anti-pattern it is possible to detect operator errors of the type "wrong order". An error message is created when a transaction event te occurs in the event stream which corresponds to an acitivity $\mathrm{a_{cur}}$, but the corresponding transaction event for the activity $\mathrm{a_{pre}}$ does not exist in the transaction event history:

$$\pi_{\mathrm{app,op,prop}}\mathrm{te} \in \pi_{\mathrm{a_{cur}.app,a_{cur}.op,a_{cur}.prop}}\mathrm{P3_i} \wedge$$
$$\pi_{\mathrm{app,op,prop}}(\pi_{\mathrm{a_{pre}}}(\sigma_{\mathrm{a_{cur}.app=te.app} \wedge \mathrm{a_{cur}.op=te.op} \wedge \mathrm{a_{cur}.prop=te.prop}}\mathrm{P3_i})) \notin$$
$$\pi_{\mathrm{app,op,prop}}\mathrm{TEH}$$

Similar pattern: The pattern IMMEDIATELY PRECEDENCE is also used to detect forgotten activities, which have to be executed before another activity. For the pattern IMMEDIATELY PRECEDENCE no activities are allowed between $\mathrm{a_{cur}}$ and $\mathrm{a_{pre}}$, whereas for the pattern PRECEDENCE addtional activites in between are allowed. Besides, the pattern is similar to the pattern NEXT. The difference is, that the pattern NEXT checks for future activities, whereas the pattern PRECEDENCE checks for activities happend in the past of a maintenance execution.

Pattern IMMEDIATELY PRECEDENCE

Description: This pattern describes which activity has to be executed immeditalely before another one. It can be used for non-concurrent activities as well as for XOR-joins and XOR-splits.

Instances: To get all instances of this pattern for a maintenance plan i we create a relation $\mathrm{P4_i} := \mathrm{AM_i}$ x $\mathrm{AM_i}$ with the tuples ($\mathrm{a_{cur}}$, $\mathrm{a_{pre}}$), where

- the corresponding transitions of the activities $\mathrm{a_{cur}}$ and $\mathrm{a_{pre}}$ are connected through the same place, and
- the corresponing transitions cannot be executed concurrent to other transitions.

Example: In figure 5 instances of the pattern IMMEDIATELY PRECEDENCE are (a2, a1), (a3, a1), (a4, a2), (a5, a3), (a6, a4) and (a7, a5).

Anti-pattern: With the anti-pattern it is possible to detect operator errors of the type “wrong order”. An error message is created when a transaction event te occurs in the event stream which corresponds to an acitivity $\mathrm{a_{cur}}$, but the latest transaction event of the same service template in the transaction event history does not correspond to $\mathrm{a_{pre}}$:

$$\pi_{\mathrm{app,op,prop}}\mathrm{te} \in \pi_{\mathrm{a_{cur}.app,a_{cur}.op,a_{cur}.prop}}\mathrm{P4_i} \wedge$$
$$\pi_{\mathrm{app,op,prop}}(\pi_{\mathrm{a_{pre}}}(\sigma_{\mathrm{a_{cur}.app=te.app} \wedge \mathrm{a_{cur}.op=te.op} \wedge \mathrm{a_{cur}.prop=te.prop}}\mathrm{P4_i})) \notin$$
$$\pi_{\mathrm{app,op,prop}}(\sigma_{\mathrm{max(timestamp)}}(\sigma_{\mathrm{timestamp} < \mathrm{te.timestamp} \wedge \mathrm{st=te.st}}))\mathrm{TEH}$$

Similar pattern: The pattern PRECEDENCE describes also activities which have to occur before another activity. In contrast to the pattern MMEDIATELY PRECEDENCE the pattern PRECENDENCE allows other activities of the same service template to occur between a pair of activities.

Pattern STATE-CONDITION

Description: This pattern describes the state an application should have in order to be able to perform an operation on either the same or another application. Example: in order to shut down an application server, the database server must be in the state offline.

Instances: Instances of this pattern are all conditions $\mathrm{SC_i}$ for a maintenance

plan i.
Example: In figure 6, which is a snippet of figure 3, the instance of the pattern STATE-CONDITION is (MyApplication, start, NULL, TestDatabase, started).

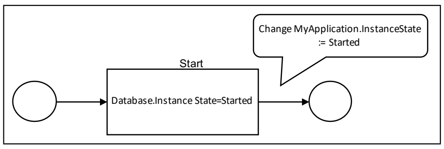

Change MyApplication.InstanceState := Started
Start
Database.Instance State=Started

Fig. 6. Example pattern STATE-CONDITION

Anti-pattern: This anti-pattern does actually not detect an error like described in chapter 2. Instead, it detects malicious prerequisites, which would lead to an operation error. This is done by comparing the latest state of an application with the planned state:

$$\pi_{\mathrm{app,op,prop}}\mathrm{te} \in \pi_{\mathrm{app,op,prop}}\mathrm{SC_i} \wedge$$
$$(\pi_{\mathrm{zapp,state}}(\sigma_{\mathrm{app=te.app}\wedge\mathrm{op=te.op}\wedge\mathrm{prop=te.prop}}\mathrm{SC_i}) \ / \ \pi_{\mathrm{app,state}}\mathrm{SEH}) \neq \emptyset$$

Similar pattern: There are no similar patterns for the STATE-CONDITION pattern.

Pattern VALUE
Description: This pattern describes the value of a configuration parameter which has to be changed during the maintenance.
Instances: To get all instances of this pattern a selection on the set of all activities of the maintenance plan is performed in order to get only those activities which include a change of a property: $\mathrm{P5} := \pi_{\mathrm{app,op,prop,value}} \ (\sigma_{\mathrm{prop} \neq \mathrm{NULL}}\mathrm{SA_i})$.
Example: In figure 3 the only instance of this pattern is (MyApplication, configure, DB Connection, jdbc:mysql://localhost:1521/XE).
Anti-pattern: This anti-pattern allows to detect operator errors of the types "wrongly executed activity", "lexical error", "local inconsistency", "global inconsistency" and "typo" by checking the element value of a transaction event te:

$$\pi_{\mathrm{app,op,prop}}\mathrm{ve} \in \pi_{\mathrm{app,op,prop}}\mathrm{P5} \wedge \pi_{\mathrm{app,op,prop,value}}\mathrm{ve} \notin \mathrm{P5_i}$$

Similar pattern: This pattern can be seen as a more detailed version of the OCCURRENCE pattern. However, the OCCURRENCE pattern just checks for executed operations and properties, but not for the actual values of modified properties.

Pattern OCCURENCE
Description: This pattern includes all activities which have to be executed in

a maintenance plan independent of the chosen path through the maintenance plan.
Instances: To get all instances of this pattern a set $\mathrm{P6_i}$ with all activities of the maintenance plan i is created, where

- every activity has to be executed independent of the chosen path in the maintenance plan

Example: In figure 3 the instances of this pattern are (MyService, MyApplication, stop, NULL, NULL), (MyService, MyApplication, configure, DB Connection, jdbc:mysql://localhost:1521/XE) and (MyService, MyApplication, start, NULL, NULL).
Anti-pattern: The anti-pattern allows to detect errors of the kind "forgotten activity". An error message is created, when an activity was executed too seldom:

$$\exists \text{ a} \in \sigma_{\text{time } > \text{ a.end}}\mathrm{P6_i} \wedge \text{countTE}(\pi_{\text{app,op,prop}}\text{a},\\ \pi_{\text{app,op,prop}}(\pi_{\text{timestamp}\geq\text{a.start } \wedge \text{ timestamp}\leq\text{a.ende}}\text{TEH})) <\\ \text{countA}(\pi_{\text{app,op,prop}}\text{a}, \pi_{\text{app,op,prop}}(\pi_{\text{start}\geq\text{a.start } \wedge \text{ end}\leq\text{a.end}}\mathrm{P6_i}))$$

Similar pattern: With the anti-patterns NEXT, IMMEDIATELY NEXT, PRECEDENCE and IMMEDIATELY PRECEDENCE it is also possible to detect forgotten activities in a limited way. However it is only possible to detect a forgotten activity right before or after another activity. As an example lets assume we have a sequence (a, b, c, d, e). If the activity a and e occur, it is possible to detect the forgotten activities b and d with the similar patterns, but not the activity c. Only with the anti-pattern OCCURENCE it is possible to detect the forgotten activity c.

Pattern ALTERNATIVE OCCURENCE

Description: This pattern describes a pair of activities which cannot be executed together, like after XOR-Splits. However, one activity of such a pair of activities has to be performed during the maintenance.
Instances: To get all instances of this pattern a relation $\mathrm{P7_i} := \mathrm{AM_i} \text{ x } \mathrm{AM_i}$ with the tuples $(\mathrm{a_{cur}}, \mathrm{a_{alt}})$, where

- the corresponding transitions of the activities $\mathrm{a_{cur}}$ and $\mathrm{a_{alt}}$ do not occur together in any path of the maintenance plan.

Example: In figure 5 instances of this pattern are (a2, a3), (a2, a5), (a2, a7), (a4, a3), (a4, a5), (a4, a7), (a6, a3), (a6, a5), (a6, a7), (a3, a2), (a3, a4), (a3, a6), (a5, a2), (a5, a4), (a5, a6), (a7, a2), (a7, a4) and (a7, a6).
Anti-pattern[1]: The anti-pattern allows to detect errors of the kind "forgotten activity". An error message is created, when activities of $\mathrm{P6_i}$ were not executed. However, no error message is created for activities, if one alternative activity was already performed. The anti-pattern assumes, that the first executed alternative activity is the right one and therefore ignores all other activities, which may not

[1] Due to space limitations we forgo the formal definition of the following anti-patterns.

be executed in conjunction with this first alternative activity.
Similar pattern: The pattern OCCURENCE does also detect forgotten activities, but it would create wrong error messages for exclusive activities, if already one of the exclusive activities was executed.

Pattern ABSENCE
Description: This pattern describes which activites may not occur during a maintenance.
Instances: The instances of this pattern are all possible activities, which could really occur during a maintenance, without all activities, which are also modelled in the maintenance plan. Note that in a maintenance plan only a subset of possible operations on service templates is modelled and therefore planned. All other operations should not occur during the maintenance. The number of instances of this pattern can get very high, because the number of possible operations, especially configurations can be huge. However, for the anti-pattern of ABSENCE the generation of pattern instances of the type ABSENCE is not needed as it is explained in the following.
Example: On the assumption that in figure 3 other possible operations of *MyApplication* would be "deploy" and "undeploy", some of the instances of the pattern ABSENCE would be (MyService, MyApplication, deploy, NULL, NULL) and (MyService, MyApplication, undeploy, NULL, NULL).
Anti-pattern: The anti-pattern detects errors of the kind "unneeded activity". An error message is created either when

- a transaction event does not correspond to one of the activities in $\mathrm{P6_i}$ or $\mathrm{P7_i}$, or
- a transaction event corresponds to one the of the activities in $\mathrm{P6_i}$ or $\mathrm{P7_i}$, but it did not occur during the planned maintenance window, or
- a transaction event corresponds to one the of the activities in $\mathrm{P6_i}$ or $\mathrm{P7_i}$ and it occured during the planned maintenance plan, but it occured too often during the maintenance window.

Similar pattern: With the anti-patterns NEXT, IMMEDIATELY NEXT, PRECEDENCE and IMMEDIATELY PRECEDENCE it is also possible to detect unneeded activities, when the following or precedence activity was not the planned one. However, these anti-patterns do not know, if the unneeded activity is an activity which was just executed in the wrong order or if it is an activity which should not occur at all.

Pattern ALTERNATIVE ABSENCE
Description: This pattern describes activities which are not allowed to be executed depending on other activities. Such activities occur after XOR-Splits.
Instances: Instances of this pattern are the same like for the pattern ALTERNATIVE OCCURENCE. Although the pattern instances are the same, the anti-pattern is different to the anti-pattern of ALTERNATIVE OCCURENCE.
Example: For an example please see the examples of the pattern ALTERNATIVE OCCURENCE.

Anti-pattern: An error message is created when one of the following conditions hold:

- A transaction event corresponds to an activity in $P7_i$. It is the first alternative activity and it occured during the maintenance window, however it was performed too often.
- A transaction event corresponds to an activity in $P7_i$ and it occured during the maintenance window, but an alternative activity was already performed before.

Similar pattern: The pattern ABSENCE is similar, but the pattern does not check the absence of activities dependent of other activities.

4.3 Derivation of pattern instances

Pattern instances can be derived from the maintenance plan by simulating it. Therefore, the maintenance plan is marked with the Service Template of the to be maintained IT Service. The resulting simulation log is used to create log-based ordering relations and footprints like they are used in process mining and described in [18], [19]. Based on these ordering relations two footprints are created. One footprint uses the basic ordering relations described in [18]. This footprint is used to derive the pattern instances IMMEDIATELY NEXT, PRECEDENCE, IMMEDIATELY PRECEDENCE and the activities a_{cur} and a_{nex} for the pattern instances of NEXT. In order to get a_{far} for the pattern instances of NEXT the second footprint is used, which is based on the extended ordering relations described in [19].
Instances of the pattern ALTERNATIVE OCCURENCE and ALTERNATIVE ABSENCE are also derived from the second footprint. The pattern OCCURENCE is instantiated with all simulated activities, which occur in every path of the simulation log. For the anti-pattern ABSENCE all possible activities are needed, which can be derived directly from the simulation log.
Instances of the pattern STATE-CONDITION can be derived from the activities in the simulation log and the corresponding function defined in definition 4.5. The pattern VALUE can be instantiated by filtering all activities in the simulation log, whose attribute *prop* is not NULL.

5 Implementation

The architecture of the proof of concept implementation consists of four main components and is shown in fig. 5. The first component is a modelling component, which allows to model maintenance plans and derive pattern instances of a maintenance plan. The modelling component is implemented in the software tool Horus[2] and already allows to model generic XML nets. The extension of the tool in order to model TOSCA service templates and link them to an XML

[2] www.horus.biz

net is currently under construction.
The second main component are the log agents. Log agents are used to get every new log entry of an application, transform the log entry into the format of a transaction event and send it to the complex event processing engine. In the proof of concept log agents are implemented with Beats and Logstash[3]. Both products are developed for fast log data extraction. Besides, Logstash contains a powerful regular expression engine, which supports the transformation of proprietary log entries into the generic format of transaction events.
The third component is an IT infrastructure monitoring tool like Nagios[4], CloudWatch[5], or Metricbeat[6] which allows to check the state of an application in order to generate the state events. In the proof of concept we use Metricbeat.
The fourth component is the complex event processing engine, which checks incoming state and transaction events against the pattern instances of the maintenance plan. In the proof of concept the complex event processing system of WSO2[7] is used. All anti-patterns are implemented as event queries in the event pattern language Siddhi[8] and have to be implemented only once. In order to check future maintenance plans, only the corresponding pattern instances have to be transferred to the complex event processing system. As an example, for an anti-pattern written in Siddhi, see the following anti-pattern NEXT, implemented as Siddhi query:

```
from te [(app == NEXT.appcur and op == NEXT.opcur
and prop == NEXT.propcur) in NEXT] insert into #temp;
from #temp as t join NEXT as n on t.app == n.appcur and
t.op == n.opcur and t.prop == n.propcur
select t.timestamp, n.appcur, n.opcur, n.propcur, n.appnex,
n.opnex, n.propnex, n.appfar, n.opfar, n.propfar
insert into #temp1;
from e1=#temp1 -> e2= incoming_te [e1.appfar == e2.app
and e1.opfar == e2.op and e1.propfar == e2.prop]
select e1.timestamp, e1.appcur, e1.opcur, e1.propcur,
e1.appnex, e1.opnex, e1.propnex, e2.timestamp as timestampfar
insert into #temp2;
from #temp2 [not((appnex == TEH.app and opnex == TEH.op
and propnex == TEH.prop in and timestamp < TEH.timestamp
and timestampentf >  TEH.timestamp) in TEH)]
select str:concat("The activity ",appnex, ", ", opnex, ", "
, propnex, " was not performed after the activity ", appcur,
" ,", opcur, ", ", propcur, ".") as message
insert into error_message;
```

[3] https://elastic.co
[4] https://nagios.org
[5] https://aws.amazon.com/en/cloudwatch/
[6] https://www.elastic.co/guide/en/beats/metricbeat/6.2/index.html
[7] https://wso2.com/products/complex-event-processor/
[8] https://github.com/wso2/siddhi

Apart of the modelling component all components and Siddhi queries are implemented in a prototype, which is used to evaluate the approach. A first evaluation experiment was conducted, which is described in the following.

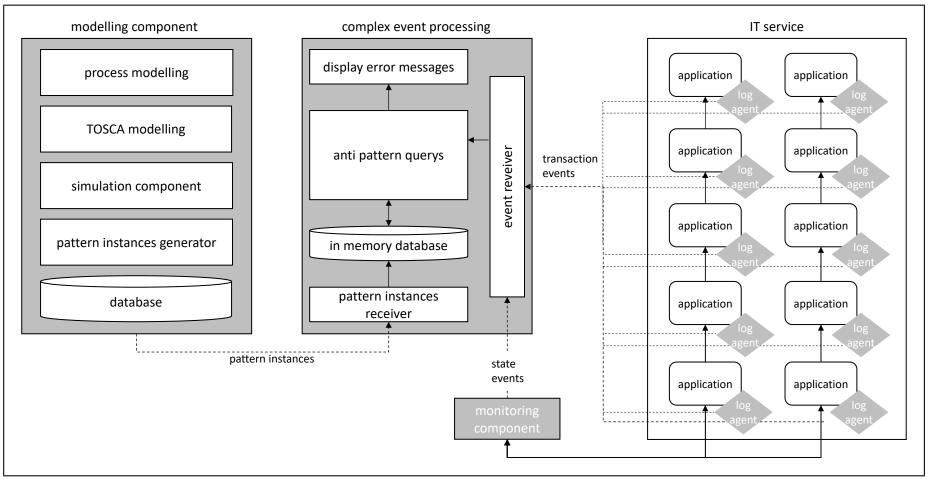


Fig. 7. Implementation architecture

6 Experimental Results

To evaluate the approach an exemplary IT service maintenance was performed. Therefore, we built an IT service environment using Amazon EC2[9] machines. The environment contains five EC2 machines. On two machines we installed an Apache webserver[10] hosting the open source application SugarCRM[11]. On two other machines Maria DB was installed. Both SugarCRM instances connect to the same Maria DB instance. In the experiment the configuration of both SugarCRM instances should be changed, so SugarCRM connects only to the second Maria DB instance anymore. The fifth EC2 machine is used to host the complex event processing engine and Logstash in order to transform log data and check the data for anti-patterns. Figure 8 gives an overview of the maintance plan, which was used for evaluation and to derive the pattern instances for the experiment. The maintenance plan is described shortly in the following.
According to the maintenance plan first of all the database server named database2 needs to be started. Afterwards the loadbalancer is shut off in order to avoid connections to the webservers. When the loadbalancer is offline, the two webservers have to be stopped and reconfigured in order to connect to database2.

[9] https://aws.amazon.com/ec2/
[10] https://httpd.apache.org
[11] https://www.sugarcrm.com

After reconfiguring and stopping the webservers, they can be started again. If both webservers where stopped, the former database server can be shut down. Finally, the loadbalancer ha to be started again in order to redirect requests to the webservers. During the execution of this maintenance plan typical operator errors, like they are described in chapter 2, were injected. Namely, those operator errors are:

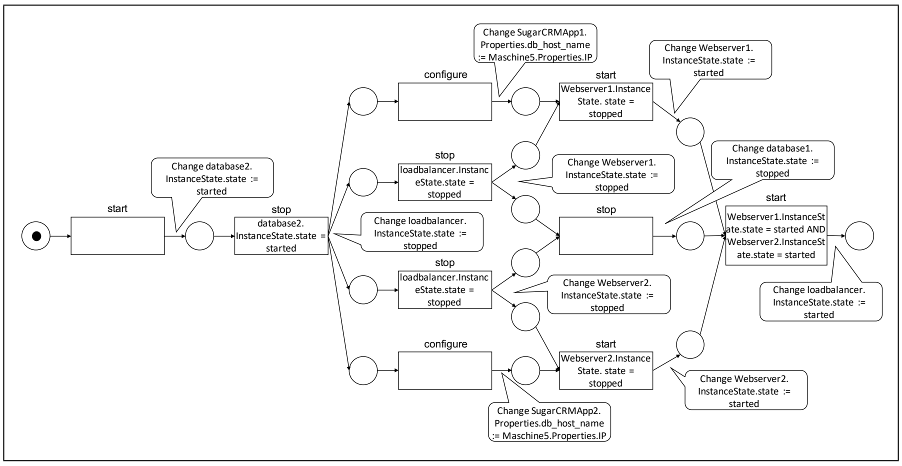

Change SugarCRMApp1. Properties.db_host_name := Maschine5.Properties.IP
Change Webserver1. InstanceState.state := started
configure
start
Webserver1.Instance State. state = stopped
Change database1. InstanceState.state := stopped
Change database2. InstanceState.state := started
stop
loadbalancer.InstanceState.state = stopped
Change Webserver1. InstanceState.state := stopped
start
stop
database2. InstanceState.state = started
Change loadbalancer. InstanceState.state := stopped
stop
start
Webserver1.InstanceState.state = started AND Webserver2.InstanceState.state = started
stop
loadbalancer.InstanceState.state = stopped
Change Webserver2. InstanceState.state := stopped
Change loadbalancer. InstanceState.state := started
configure
start
Webserver2.Instance State. state = stopped
Change Webserver2. InstanceState.state := started
Change SugarCRMApp2. Properties.db_host_name := Maschine5.Properties.IP

Fig. 8. Maintenance plan used for the experiment

1. Forgot to configure SugarCRM1
2. The loadbalancer was started before webserver2 was started
3. When configuring SugarCRM1 the IP address is changed and additionally without need the user is modified
4. Instead of stopping databaseserver1, databaseserver2 is stopped
5. The wrong IP address is entered, when configuring SugarCRM1
6. When changing the IP address a typo happens, so that the format of the IP address is xxx.xxx.xxxxxx instead of xxx.xxx.xxx.xxx
7. Webserver2 is started. However, the EC2 machine is stopped manually in order to simulate a software bug which hinders the webserver to start properly
8. A combination of error 1. and 2.
9. A combination of error 3. and 4.
10. A combination of error 3., 4. and 6.

In the first ten runs of the experiment one error was injected per run. Afterwards we repeated the experiment. However, in the second ten runs, errors were corrected immediately after their detection. By correcting them, the actual maintenance execution differs from the maintenance plan, because we did not model any procedural exception handlings.

In order to quantify the results, we used the metrics *Recall*, *Precision* and *F-Score* known from machine learning evaluations [24].

In the first ten runs all errors were identified and no false positives were reported, resulting in a Precision and Recall of 100%. The negative site of this is, that error messages were created multiple times for the same root cause. For example when the configuration of SugarCRM1 was forgotten (error 1) the anti-pattern NEXT as well as the anti-pattern PRECEDENCE created an error message, when the activity after the forgotten acitivity was executed. Additionally, the anti-pattern OCCURENCE created an error message, because the occurence of the activity could not be found in the transaction event history. In some runs this led to a ratio of up to four created error messages for one root cause.

In the second round of the experiment, when the errors were corrected after their identification, the precision decreased to 41%. The reason was the increasing number of false positives. For example when the forgotten configuration of SugarCRM1 was identified, the error was corrected by stopping webserver1 again, configuring SugarCRM1 and starting webserver1 again. When webserver1 was stopped respectively started the anti-pattern ABSENCE created two error messages that webserver1 was stopped respectively started too many times. Nonetheless, all injected errors were identified resulting in a Recall of 100%. Besides, the ratio of reported error messages to root causes improved, because of the immediate error handling. The F-Score over all runs is 73%. The F-Score of the first ten runs is 100%, whereas the F-Score only for the runs with immediate error handling is 58%.

In the next months we plan to conduct additional experiments to test the recall and precision of the method. Besides, we plan to perform performance tests as this is one of our main objectives, that the method reports errors within seconds. So, operators would have a realistic chance to correct their errors before they manifest in an IT service.

7 Discussion

The high false positive rate when error handling was performed can be interpreted as overfitting. On the contrary, the false positive rate would decrease, if error handling would be modelled explicitly in the maintenance plan. However, we think it is quite unlikely that an operator would model every possible exception, because this could lead to a very complex unreadable maintenance plan.

In general checking conformance of an event stream to a process model in an online setting leads to new challenges, that do not exist for traditional offline conformance checking methods. In offline conformance checking methods, analyzing event traces instead of event streams, the actual process execution can not be influenced anymore (apart of long running process, which are not finished when analyzing the log trace). In an online setting like in this work an operator would adjust the process execution spontaneous, because of the idenfitication of errors or deviations from the process modell. This could lead to process executions differing a lot from the actual modelled process.

An error detection method or conformance checking method should be able to recognize, which deviations from the process modell are allowed ones, e.g. because of correcting an error and which one are real errors in order to reduce false positives. Therefore, we plan to extend our approach by adding a machine learning component, which analyzes which error messages are real errors and which were created because of correcting an already performed error.

8 Related Work

Related work can be separated in different areas of work. One area of work is the automation of typical operations like redeployments and integrated error exception handling, like it is provided by popular configuration management tools, e.g. Chef [21]. Those tools have the disadvantage, that they have just local information for error handling and no global view of the whole maintenance, which also could involve legacy systems [20].

Another area of work is the detection of configuration errors. Those approaches can be divided in rule based methods and online configuration validation [22]. Rule based methods try to avoid configuration errors a priori by correctness checks. These, help to detect wrong planned configuration errors. However, those approaches do not check if the configuration operation itself was executed as planned. So, forgotten configurations e.g. because a server was down or typos, when the configuration was done manually, cannot be detected.

The most related work to ours is the work of Xu et al. [20] and Farshchi et al. [23]. Both works describe an approach to monitor sporadic operations in cloud environments. Xu et al. developed a method called "POD-Diagnosis". They use a process model to detect operator errors through token replay by checking the conformance of observed logs with the prebuild model and an additional fault tree analysis in order to find the root cause of the error. In contrast to our work only the control flow of the process is modelled and can therefore be checked. Apart of that, in our approach no additional fault tree has to be build. Farshchi et al. build a regression-based model to find correlation and causalities between events described in logs and overserved metrics of resources. In their approach, assertions are derived from the regression-based model. However, they are also limited to control flow. Additionally, enough learning data is needed, which practically limits their approach to automated cloud environments. Our approach does not have to learn data and therefore can also be used to monitor manually executed steps or changes in legacy systems as long as those actions can be seen in the logs of the systems.

The field of business process compliance monitoring can also be seen as other related work, which uses patterns to check that process executions adhere to predefined compliance requirements [26]. Our work differs from this field in the way, that we do not use patterns to make sure, that specific compliance requirements like the "segregation of duty" pattern are covered during a process execution. The patterns used in this work are only used in order to be able to instantiate anti-patterns during process executions in order to identify situations, which must not occur. The notion of anti-pattern can be seen as the counterpart to compliance pattern [26]. The most related work to ours in this field is [27], who also use the notion of anti-patterns instead of compliance patterns. However, in our work no definition of patterns has to be perfomed by an user. Instances of the patterns are derived automatically from the created maintenance plan during design phase, which integrates the control flow and configuration modelling. The instantiation of patterns can even be influenced after modelling by the simulation of the maintenance plan. If paths of the maintenance plan will not be simulated, those will not be represented in the simulation log and therefore cannot be used for pattern instantiation. During the execution phase anti-patterns check only deviations from the derived pattern instances.

9 Conclusion

In this paper, we describe an approach to detect operator errors online during the execution of maintenance operations. Therefore, we define different anti-patterns, which are implemented as complex event processing queries and check in real time log entries and state metrics of observed resources against pattern instances of a predefined process model. The process model itself is realized as a TOSCA based XML net, combining the modelling of the control-flow and the resources. A first evaluation with a prototype implementation was performed, resulting in a very good error detection rate. On the contradictory site, the approach can result in a high false postive rate, when the process execution is adjusted spontaneously in order to correct the reported errors. Therefore, we plan to extend our approach in order to deal with this spontaneous flexibility during an IT service maintenance.

References

1. H. S. Gunawi et al., "What Bugs Live in the Cloud?: A Study of 3000+ Issues in Cloud Systems," in Proceedings of the ACM Symposium on Cloud Computing, pp. 1-14 (2014)
2. S. Hagen, M. Seibold, and A. Kemper, "Efficient verification of IT change operations or: How we could have prevented Amazon's cloud outage," presented at the Network Operations and Management Symposium (NOMS), 2012 IEEE, pp. 368-376 (2012)
3. T. Dumitra? and P. Narasimhan, "Why do upgrades fail and what can we do about it?: toward dependable, online upgrades in enterprise system," in Proceedings of the 10th ACM/IFIP/USENIX International Conference on Middleware, p. 18. (2009)
4. S. Pertet and P. Narasimhan, "Causes of failure in web applications," Parallel Data Lab., p. 48, (2005)
5. D. Oppenheimer, A. Ganapathi, and D. A. Patterson, "Why Do Internet Services Fail, and What Can Be Done About It?," in Proceedings of the 4th Conference on USENIX Symposium on Internet Technologies and Systems - Volume 4, Berkeley, CA, USA, pp. 1-1 (2003)
6. D. Scott, "Making smart investments to reduce unplanned downtime," Tactical Guidel. Res. Note Note TG-07-4033 Gart. Group Stamford CT (1999)

7. S. Elliot, "DevOps and the cost of downtime: Fortune 1000 best practice metrics quantified," Int. Data Corp. IDC (2014)
8. A. Vetter, "Detecting Operator Errors in Cloud Maintenance Operations," in 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), pp. 639-644 (2016)
9. K. Nagaraja, F. Oliveira, R. Bianchini, R. P. Martin, and T. D. Nguyen, "Understanding and Dealing with Operator Mistakes in Internet Services," in OSDI '04: 6th Symposium on Operating Systems Design and Implementation (2004)
10. Z. Yin, X. Ma, J. Zheng, Y. Zhou, L. N. Bairavasundaram, and S. Pasupathy, "An empirical study on configuration errors in commercial and open source systems," in Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, pp. 159-172 (2011)
11. J. L. Peterson, "Petri net theory and the modeling of systems," (1981)
12. E. Gamma, R. Helm, R. Johnson, and J. Vlissides, Design patterns: elements of reusable object-oriented software. Pearson Education (1994)
13. W. M. van der Aalst, A. H. Ter Hofstede, B. Kiepuszewski, and A. P. Barros, "Workflow patterns," Distrib. Parallel Databases, vol. 14, no. 1, pp. 5-51 (2003)
14. N. Russell, A. H. Ter Hofstede, D. Edmond, and W. M. van der Aalst, "Workflow data patterns," QUT Technical report, FIT-TR-2004-01, Queensland University of Technology, Brisbane (2004)
15. N. Russell, A. H. Ter Hofstede, D. Edmond, and W. M. van der Aalst, "Workflow resource patterns" (2005)
16. D. Riehle and H. Zllighoven, "Understanding and using patterns in software development," TAPOS, vol. 2, no. 1, pp. 3-13 (1996)
17. M. B. Dwyer, G. S. Avrunin, and J. C. Corbett, "Property specification patterns for finite-state verification," in Proceedings of the second workshop on Formal methods in software practice, pp. 7-15 (1998)
18. W. Van Der Aalst, Process mining: discovery, conformance and enhancement of business processes. Springer Science Business Media (2011)
19. M. Weidlich, J. Mendling, and M. Weske, "Computation of behavioural profiles of process models," Bus. Process Technol. Hasso Plattner Inst. IT-Syst. Eng. Potsdam (2009)
20. X. Xu, L. Zhu, I. Weber, L. Bass, and others, "POD-diagnosis: Error diagnosis of sporadic operations on cloud applications," in 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 252-263 (2014)
21. Chef, "About Handlers," 08-Nov-2017. [Online]. Available: https://docs.chef.io/handlers.html.
22. T. XU and Y. ZHOU, "Systems Approaches to Tackling Configuration Errors: A Survey" (2014)
23. M. Farshchi, J.-G. Schneider, I. Weber, and J. Grundy, "Metric selection and anomaly detection for cloud operations using log and metric correlation analysis," J. Syst. Softw. (2017)
24. D. Powers, "Evaluation: From precision, recall and f-measure to roc., informedness, markedness correlation" J. Mach. Learn. Technol., vol. 2, no. 1, pp. 37-63, 2011
25. O. Kopp, T. Binz, U. Breitenbcher, and F. Leymann, "BPMN4TOSCA: A Domain-Specific Language to Model Management Plans for Composite Applications" in Business Process Model and Notation, J. Mendling and M. Weidlich, Eds. Springer Berlin Heidelberg, 2012, pp. 38-52
26. M. Becker and S. Klingner, "A Criteria Catalogue for Evaluating Business Process Pattern Approaches" in Enterprise, Business-Process and Information Systems

Modeling, I. Bider, K. Gaaloul, J. Krogstie, S. Nurcan, H.A. Proper, R. Schmidt, P. Soffer, Eds. Springer Berlin Heidelberg, 2014, pp. 257-271

27. A. Awad, A. Barnawi, A. Elgammal, R. Elshawi, A. Almalaise, and S. Sakr, "Runtime Detection of Business Process Compliance Violations: An Approach Based on Anti Patterns", 12th Enterp. Eng. Track Acm Sac15, 2015.