Model-Driven Trace Diagnostics for Pattern-based Temporal Specifications

Wei Dou  
University of Luxembourg  
dou@svv.lu

Domenico Bianculli  
University of Luxembourg  
domenico.bianculli@uni.lu

Lionel Briand  
University of Luxembourg  
lionel.briand@uni.lu

ABSTRACT
Offline trace checking tools check whether a specification holds on a log of events recorded at run time; they yield a verification verdict (typically a boolean value) when the checking process ends. When the verdict is false, a software engineer needs to diagnose the property violations found in the trace in order to understand their cause and, if needed, decide for corrective actions to be performed on the system. However, a boolean verdict may not be informative enough to perform trace diagnostics, since it does not provide any useful information about the cause of the violation and because a property can be violated for multiple reasons.

The goal of this paper is to provide a practical and scalable solution to solve the trace diagnostics problem, in the settings of model-driven trace checking of temporal properties expressed in TempPsy, a pattern-based specification language. The main contributions of the paper are: a model-driven approach for trace diagnostics of pattern-based temporal properties expressed in TempPsy, which relies on the evaluation of OCL queries on an instance of a trace meta-model; the implementation of this trace diagnostics procedure in the TemPsy-Report tool; the evaluation of the scalability of TemPsy-Report, when used for the diagnostics of violations of real properties derived from a case study of our industrial partner. The results show that TemPsy-Report is able to collect diagnostic information from large traces (with one million events) in less than ten seconds; TemPsy-Report scales linearly with respect to the length of the trace and keeps approximately constant performance as the number of violations increases.

CSC CONCEPTS
• Software and its engineering → Software verification and validation;

KEYWORDS
OCL, temporal constraints, trace diagnostics, offline trace checking, run-time verification, specification patterns

1 INTRODUCTION
Run-time verification (RV) [31] is a verification technique used for checking the correctness of an execution of a system with respect to a specification. The checking procedure (represented by a monitor program) can be performed during the actual system execution (a settings called online monitoring) or by analyzing a log of recorded events produced by the system (a settings called offline monitoring or offline trace checking).

Among the many approaches for run-time verification [4], our previous work on model-driven trace checking [17]—developed as part of a research project in collaboration with our public-service partner CTIE (Centre des technologies de l’information de l’Etat, the Luxembourg national center for information technology)—was focused on model-driven run-time verification of business processes [16]. Our approach checks properties expressed in TempPsy (Temporal Properties made easy) [11, 17], a pattern-based domain-specific language for the specification of temporal requirements. The approach, implemented in the TemPsy-Check tool [18], relies on an optimized mapping of temporal requirements written in TempPsy into Object Constraint Language (OCL) constraints on a meta-model of execution traces. More specifically, it reduces the problem of checking a temporal property over an execution trace to the evaluation of an OCL constraint (derived from the property to check and semantically equivalent to it) on an instance of the trace meta-model.

Run-time verification tools (more specifically, monitors) yield a verification verdict (for short, verdict) after checking a property over an execution trace: the verdict is a truth value from some truth domain [31]. In the case of offline trace checking, typically the verdict is boolean (true/false). When the verdict is false, a software engineer needs to diagnose the property violations found in the trace in order to understand their cause and, if needed, decide for corrective actions to be performed on the system. However, a boolean verdict may not be informative enough to perform trace diagnostics. For example, let us consider the following temporal property, informally stated in English: “it is always the case that if event A occurs then it should stimulate, within 5 time units, the sequential occurrence of events B followed, within 2 time units, by C”; it corresponds to a time-constrained response chain pattern [2]. This property can be violated for various reasons, such as:

• there is at least an occurrence of A not followed by the sequence of events B-C;
• there is at least an occurrence of A that is followed by the...
The technique follows a model-driven approach, based on the trace violated in different ways. Just yielding a boolean verdict for reporting a violation in the support a better understanding of violations. Furthermore, we have developed an interactive visualization tool that provides a graphical visualization of the diagnostic information collected by TemPsy-Report, to support a better understanding of violations.

The trace diagnostics technique proposed in this paper leverages the pattern-based, flat (i.e., without operator nesting) structure of temporal properties expressed in TemPsy for the precise characterization of the violations that can occur with each type of property. Based on this characterization, the key idea is to retrieve, from a trace that violates a property, the diagnostic information that describes the violations admitted by the specific type of property. The technique follows a model-driven approach, based on the trace meta-model defined in our previous work [17]: it uses OCL queries defined on the trace model to analyze violations and collect diagnostic information. The queries are supported by a set of auxiliary OCL functions and optimized based on the structure of the targeted TemPsy property for achieving better performance.

We evaluated the scalability of our TemPsy-Report tool by studying how the execution time varies depending on the trace length, the number of violations to diagnose, and the violation and property type. We used a benchmark of TemPsy properties based on real requirements extracted from our case study, on traces with length ranging from 100K to 1M. The results show that TemPsy-Report is able to collect diagnostic information from large traces (with one million events) in less than ten seconds; it scales linearly with respect to the length of the trace and keeps approximately constant performance as the number of violations increases.

To summarize, the main contributions of this paper are: i) a model-driven approach for trace diagnostics of pattern-based temporal properties expressed in TemPsy, which relies on the evaluation of OCL queries on an instance of a trace meta-model; ii) a publicly available tool TemPsy-Report, implementing the trace diagnostics procedure; iii) an evaluation of the scalability of TemPsy-Report, when applied to the diagnostics of violations of real properties derived from a case study of a complex information system developed and used by the Luxembourg government.

We have implemented the approach proposed in this paper in a publicly available tool [12, 13]. Furthermore, we have developed an interactive visualization tool that provides a graphical visualization of the diagnostic information collected by TemPsy-Report, to support a better understanding of violations.

The rest of the paper is structured as follows. Section 2 provides some background information about the TemPsy language. Section 3 presents a classification of violations for the patterns supported by TemPsy. Section 4 illustrates TemPsy-Report, our model-driven procedure for trace diagnostics. Section 5 describes tool support. Section 6 reports on the evaluation of the scalability of the TemPsy-Report tool. Section 7 discusses related work. Section 8 concludes the paper and gives directions for future work.

2 BACKGROUND: THE TEMPSY LANGUAGE

TemPsy [11, 17] is a pattern-based, domain-specific language for the specification of temporal properties. It has been developed based on the analysis of the requirements of various applications implementing business process models in the context of eGovernment systems [16]. The analysis established that all the requirements of the case study could be expressed as temporal properties by using the property specification patterns proposed by Dwyeer et al. [19], with some additional expressions. The resulting language supports the eight patterns ("absence", "universality", "existence", "bounded existence", "precedence", "response", "precedence chain", "response chain") and the five scopes ("globally", "before", "after", "between-and", "after-until") defined in [19]; patterns represent high-level abstractions of formal specifications while scopes indicate the portions of a system execution in which a certain pattern should hold. The new extensions introduced by TemPsy are: (1) the possibility, in the definition of a scope boundary, to refer to a specific occurrence of an event; (2) the possibility to indicate a time distance with respect to a scope boundary; (3) support for expressing time distance between occurrences in the precedence and response patterns (hereafter collectively called order patterns) as well as their chain versions; (4) additional variants for the bounded existence and absence patterns.

Space reasons, we only explain TemPsy informally; we refer the reader to the extended version [15] of our previous work [17] for a more complete and formal treatment. A TemPsy property includes two main entities: a scope and a pattern; both are denoted by keywords with an intuitive syntax. Properties cannot be nested.
Events used in properties are alphanumeric strings, matching the event names logged in the execution trace on which the properties specified in TemPsy are meant to be checked. TemPsy properties may contain time distances (both between events and from scope boundaries); time distances are expressed with an integer value, followed by the ‘tu’ keyword, which represents a generic system time unit (i.e., any denomination of time) as suggested in [30]). Chains of events, used in order patterns, are defined as comma-separated list of events, possibly with a time distance between each pair of events (denoted with the ‘#’ symbol).

As an example, the property “Event B shall happen at least 4 time units before the third occurrence of event Y” is expressed in TemPsy as “before 3 Y at least 4 tu eventually B”.

The semantics of patterns in TemPsy is defined as follows:

**Universality.** An event should occur across the entire execution trace; the corresponding keyword is always.

**Existence.** The existence pattern can be expressed in four variants, using the following syntax: “eventually [at least | at most | exactly | exactly]

**Absence.** In addition to stating that a certain event never occurs in the given scope, TemPsy makes also possible to specify that a specific number of occurrences of the same event should not happen, as in “never exactly 2 A”, which indicates that A should never occur exactly twice.

**Precedence.** This pattern indicates the precondition relationship between a pair of events (respectively, the two blocks of a chain). A block of an event chain can be either an atomic event or a sequence of individual events with optional constraints on the time distance between two consecutive events within the block. In this pattern, the occurrence of the second event (respectively, block) depends on the occurrence of the first event (respectively, block). Based on this definition, we added support for timing information to enable expressing the time distance between two adjacent events. For example, the pattern “A preceding at most 5 tu B, #at least 2 tu C” indicates that the event A is the precondition of the block “B followed by C”. In this pattern, A (left-hand side of ‘preceding’) represents the first block, while the expression “B, #at least 2 tu C” represents the second block. The time distance between the two blocks, specified right after ‘preceding’ and equivalent to the distance between the two timestamps of the second block (i.e., B) and the last element of the first block (i.e., A), should be at most 5 (time units). The time distance between a specific pair of consecutive events in the same block (denoted with a # symbol), in this case B and C, should be at least 2.

**Response.** This pattern specifies the cause-effect relationship between a pair of events (respectively, the two blocks of a chain) in which the occurrence of the first event (respectively, block) leads to the occurrence of the second event (respectively, block). Similarly to precedence, we added support for timing information to enable expressing the time distance between two adjacent events.

### 3 Characterization of TemPsy Violations

At the basis of our model-driven approach for trace diagnostics there is a precise characterization of the violations that can occur with each type of property. Since TemPsy does not allow for arbitrary nesting of temporal expressions (i.e., the structure of temporal properties in TemPsy is flat), such a characterization yields a finite set of possible violations, which can be uniquely associated with each type of property. In the following, we describe the different violation types that characterize TemPsy properties. In the examples, we represent an execution trace as a list (denoted with brackets) of trace elements; each trace element is a pair (denoted with parentheses) consisting of the event name and a timestamp.

**UNOC Unexpected Occurrence.** This type of violation is triggered by unexpected occurrences of the event specified in an absence or existence pattern.

The original version of the absence pattern, by definition, is violated by any occurrence of the event specified in the pattern. The variant of the absence pattern having exactly as comparison operator triggers this violation type when the number of event occurrences in the trace is equals to the number specified in the pattern. As for the existence pattern, the two variants with at most/exactly in the constraint on the number of occurrences are violated when the bound is exceeded because of an unexpected occurrence. For instance, given the trace \([A, 2], (A, 3), (A, 5)\] and the TemPsy property “globally eventually at most 2 A”, the third trace element is unexpected with respect to the existence pattern (which in this case sets a bound of at most two occurrences of A) and hence triggers a UNOC violation.

**NSOC No-Show Occurrence.** This type of violation is the dual of UNOC: it is triggered upon detecting a missing occurrence of the event specified in a universality or existence pattern.

In the case of a universality pattern, an NSOC violation is triggered whenever a trace element does not match the event specified in the pattern.

In the case of an existence pattern, the basic variant and the ones with exactly/at least in the constraint on the number of occurrences are violated when the actual number of occurrences of the event is less than the lower bound specified in the property. For example, the TemPsy property “globally eventually at least 2 A”, when checked on the trace \([A, 2], (B, 3), (B, 5)\], yields an NSOC violation because the number of occurrences of event A is less than two.

**NSOR No-Show Order.** This type of violation is triggered while checking an order pattern, when one of the two blocks of events does not occur according to the order defined by the pattern. For instance, the TemPsy property “globally A preceding B”, when checked on the trace \([b, 2], (a, 3), (c, 5)\], gives an NSOR violation because there is no occurrence of event A that precedes an occurrence of event B.

**WTO Wrong Temporal Order.** This type of violation is specific to an order pattern that contains a constraint on the time distance between the two blocks of the pattern; it is triggered when this constraint is violated. For example, the trace \([A, 2], (B, 6), (C, 15)\] violates the property “globally A, #at least 3 tu B preceding at most 2 tu C” and yields a WTO because the distance between
We recall that our meta-model for traces contains a class Trace which is composed of a sequence of words; this violation combines (and supersedes) the violations when both occur within the same property.

WTOC Wrong Temporal Order and Chain. This type of violation is specific to an order pattern that contains at least one constraint on the distance between two consecutive events in one of the two blocks; it is triggered when such a constraint is violated. For instance, the trace $[(A, 2), (B, 3), (C, 5)]$ violates the property "globally A, at least 3 tu B preceding at most 2 tu C" because, though C is preceded by the event chain A, B, the distance between event A and B (3-2=1 time unit in this case) is less than the prescribed lower bound (3 time units); a WTOC violation is then triggered.

4 MODEL-DRIVEN TRACE DIAGNOSTICS

Our technique for trace diagnostics is based on the idea of retrieving, from a trace $\lambda$ that violates a property $\rho$, the diagnostic information that is associated with the possible violations admitted by $\rho$, as determined by the characterization of TemPsy violations presented in the previous section. Notice that we expect the trace diagnostics to be run after performing trace checking (for example, by means of TemPsy-Check [18]): this is why we assume that $\lambda$ is known to violate $\rho$.

The technique follows a model-driven approach, to fulfill requirement R1 stated in Section 1. More specifically, we use the meta-model for traces introduced in our previous work [17]; we define OCL queries on this meta-model to analyze violations and collect diagnostic information. The queries are supported by a set of auxiliary OCL functions, which are optimized based on the structure of the targeted TemPsy property for achieving better performance when processing large traces, with a huge number of events.

4.1 Overview of the approach

We recall that our meta-model for traces contains a class Trace, which is composed of a sequence of TraceElements; each TraceElement is represented by a pair (event, timestamp), whose elements correspond to the actual event recorded in the trace and the time at which it occurred. One of the attributes of class Trace is properties, which is a collection of TemPsyExpressions, representing the TemPsy properties to analyze for trace diagnostics. More details on this meta-model are available in [17].

The approach takes as input a log file (i.e., a trace) and a set of TemPsy properties for which to compute the diagnostics; the input trace is converted to an instance of class Trace, while the properties are converted to instances of the TemPsyExpression class.

At the core of our approach there is the evaluation of OCL queries on the instance of the Trace class, to collect diagnostic information from the trace instance based on the TemPsy properties to analyze. These queries follow the template shown in Figure 1: for a given TemPsy property provided in input, they apply the semantics of the pattern used in the input property on a set of sub-traces, as defined by the scope used in the property. Through the application of the pattern semantics, the query collects the relevant diagnostic information, specific to the type(s) of violations found. More in detail, the expression at lines 6–15 iterates through each sub-trace to collect diagnostic information. In each iteration, if a violation is found in the sub-trace, a new triple that contains the diagnostic information is collected (and appended to the sequence result, see lines 11–14). Each triple consists of two integers (begin and end), which indicate the boundaries of each trace segment, and the diagnostic information for the violation (violations).

Figure 1: The template for OCL queries on a trace for collecting diagnostic information
Algorithm 1: diagInfExistence

\begin{algorithm}
\textbf{Input:} begin, end: the boundaries of a sub-trace; pattern: an instance of the existence pattern of the form \text{"eventually [op n] E"}
\textbf{Output:} result: a pair which contains a violation type and a list of locations related to the violation
\begin{algorithmic}
\State $E \leftarrow$ event name in pattern
\State $op \leftarrow$ comparison operator of the bound on the number of occurrences of event $E$
\State $n \leftarrow$ threshold of the number of occurrences of event $E$
\State $result \leftarrow$ null, $locations \leftarrow []$
\For {$i \leftarrow$ begin to end do}
\If {self.traceElements[i] $=$ $E$}
\State $locations \leftarrow$ append($i$)
\EndIf
\EndFor
\State $count \leftarrow$ locations.size()
\If {$op \neq$ null then}
\If {$count < n \&\& op \neq \text{"at most"}$}
\State $result \leftarrow$ (NSOC, locations)
\ElseIf {$count > n \&\& op \neq \text{"at least"}$}
\State $result \leftarrow$ (UNOC, locations[$n+1$, count])
\ElseIf {$count = 0$}
\State $result \leftarrow$ (NSOC, [])
\EndIf
\EndIf
\EndIf
\State return result
\end{algorithmic}
\end{algorithm}

input the two boundaries of a sub-trace and the pattern used in the property (accessed through the expression property, pattern) and returns the diagnostic information about the violations found in the sub-trace for the pattern given in input. There are five functions to substitute for this placeholder: diagInfUniversality, diagInfExistence, diagInfAbsence, diagInfPrecedence, and diagInfResponse. The complete definitions in OCL of these functions are available in the first author’s PhD thesis [11, Chapter 4]. For space reasons, in the following subsections we illustrate only two of these functions; for readability and conciseness, all the code snippets presented are written using OCL pseudocode.

### 4.2 Diagnostics for the “Existence” Pattern

The function for collecting diagnostic information for the existence pattern is shown in Algorithm 1. The function takes as input the boundaries of a sub-trace and an instance of the existence pattern of the form \text{"eventually [op n] E"}, and returns a pair that contains the type of violation and a list of events with locations related to the violation. After reading the parameters $E$, $op$, and $n$ from the instance of the existence pattern (lines 1–3), the function initializes the variable $result$ (storing the output value) to null and the auxiliary variable $locations$ to an empty list. The latter is then populated with all the locations in the sub-trace in which event $E$ occurs (loop at lines 5–7). Then, depending on the threshold $n$ of the bound on the number of occurrences of event $E$, the function determines the violation type and the occurrences that trigger the violation (lines 9–15).

As described in Section 3, when an existence pattern is violated, it can yield either a UNOC or NSOC violation type, depending on the variant of the pattern. If the variant of the existence pattern has a comparison operator, then the procedure checks which violation should be triggered. More precisely, if the comparison operator is either \text{"at least"} or \text{"exactly"} (line 10) and the number of occurrences of $E$ is less than $n$, then the violation type will be NSOC and the list of locations will contain all the locations in which event $E$ occurs (line 11); the latter are selected because they represent the “fragment” in the sub-trace in which the missing occurrence(s) of $E$ was supposed to appear. Otherwise, if the number of occurrences of event $E$ is more than $n$ and the comparison operator is either \text{"at most"} or \text{"exactly"} (line 12), then the violation type will be UNOC and the list of locations will contain all the locations with the “extra” occurrences of event $E$, i.e., all the locations after the $n$-th occurrence of event $E$ (line 13). If the existence pattern is in its basic variant (i.e., with no explicit constraint on the number of occurrences), if there is no occurrence of event $E$ in the sub-trace, then the violation type is set to NSOC, with an empty list of locations related to the violation (line 14). The function ends by returning the value of variable $result$.

### 4.3 Diagnostics for the “Precedence” Pattern

Function diagInfPrecedence (Algorithm 2) defines the algorithm for the variant of the precedence pattern that contains no time constraint on the distance between the two blocks. The function takes as input the two boundaries of a sub-trace and an instance of the precedence pattern of the form \text{"block1 preceding block2"}. We recall that both blocks of the pattern can be either an atomic event or a chain of events with optional constraints on the time distance between two consecutive events within the block. The function returns a list of triples, each of which contains a violation type, the location of the occurrence of block2, and the corresponding location of block1 related to the violation, and the location of the corresponding occurrence of block1.

After reading block1 and block2 from the instance of the precedence pattern (line 1), the function initializes (lines 2–6) variable $result$ (storing the return value) to an empty list and some auxiliary variables: $size_1$ and $size_2$ store, respectively, the size of block1 and of block2; firstOfBlock1 and firstOfBlock2 contain the first event defined, respectively, in block1 and block2; the tuple $(i_1, t_1)$ and $(i_2, t_2)$ are used to track whether the trace element being matched is part of an occurrence of block1 (respectively, block2); the first element of the tuple stores the position within the block of the next event to be matched while the second tuple element stores the timestamp of the previous trace element matched at position block1[$i_1$] and block2[$i_2$], respectively; variable flag is used to track whether the ongoing match of block1 is consistent with the distance constraints (if defined) within the block; variable last contains the location of the last occurrence of block1 that was found to violate the time constraints within the block.

As presented in Section 3, when a precedence pattern with no time constraint on the distance between the two blocks is violated, it can yield either a NSOR or a WTC violation type; the latter case can occur only if block1 contains a distance constraint1 that is violated (represented by block1).

---

1Based on the semantics of the precedence pattern [11], if block1 contains a distance constraint that is violated it will not be matched, i.e., it will not be considered an "effect" that must be preceded by a "cause" (represented by block1).
Algorithm 2: diaglnPrecedence

Input: begin, end: the boundaries of a sub-trace; pattern: an instance of the precedence pattern of the form “block₁ preceding block₂.”

Output: result: a list of triples, each of which consists of a violation type, the location of the occurrence of block₂ related to the violation, and the location of the corresponding occurrence of block₁

1. block₁, block₂ ← read the blocks from pattern
2. result ← []
3. size₁, size₂ ← the sizes of block₁ and block₂
4. firstOfBlock₁ ← block₁.first().event, firstOfBlock₂ ← block₂.first().event
5. \((i₁, t₁) ← (1, 0), (i₂, t₂) ← (1, 0)\)
6. flag ← true, \(l_w ← 0\)
7. for \(i ← \text{begin to end} \) do
8.     \(\text{elem} ← \text{self.traceElements}[i]\),
9.     \((e, t) ← (\text{elem}.event, \text{elem}.timestamp)\)
10.    if \(e = \text{firstOfBlock₁}\) then \((i₁, t₁, \text{flag}) ← (2, t, \text{true})\)
11.    else if \(i₁ > 1\) then \((i₁, t₁, \text{flag}) ← \text{matchSecondaryBlock}(\text{block₁}(i₁, t₁, \text{flag}), (e, t))\)
12.    if \(l_w > 0 \&\& \text{flag}\) then \((i₂, t₂) ← (1, 0)\)
13.    else if \(i₂ = \text{size} + 1\) then
14.        if \(\text{flag}\) then break
15.        else
16.            \(l_w ← i\), \((i₁, t₁) ← (1, 0), (i₂, t₂) ← (1, 0)\)
17.    if \(e = \text{firstOfBlock₂}\) then \((i₂, t₂) ← (2, t)\)
18.    else if \(i₂ > 1\) then \((i₂, t₂) ← \text{match}(\text{block₂}(i₂, t₂, (e, t))\)
19.    if \(i₂ = \text{size} + 1\) then
20.        if \(l_w = 0\) then \(v ← (\text{NSOR}, i, \text{null})\)
21.        else \(v ← (\text{WTC}, i, l_w)\)
22.        result.append(\(v\))
23.        \((i₁, t₁) ← (1, 0), (i₂, t₂) ← (1, 0)\)
24. return result

otherwise. At line 11, the function continues to check whether the matched event is part of an invalid occurrence of block₁. If there is already an invalid occurrence of block₁ (i.e., \(l_w > 0\)) and the matched event violates the constraint on the distance between block₁ \([i₁ - 1]\) and block₁ \([i₁]\), the algorithm resets the tuple \((i₁, t₁)\) (line 11). If it is not the case and block₁ is fully matched, if variable \(\text{flag}\) is true, the function stops the collection of diagnostic information (line 13); otherwise, the variable \(l_w\) is updated with the position \(i\) of the current invalid occurrence of block₁ (line 15) and the variables \(i₁, t₁, i₂, t₂\) are reset (line 16).

As an example, given the trace \([A, 2], (B, 6), (C, 15)\), \(\text{begin} = 1\), \(\text{end} = 3\), and (a property with) the precedence pattern “A, #at least 3 tu B preceding C”, the function will execute the following main steps: 1) when \(i = 1\), \((i₁, t₁, \text{flag})\) is set to \((2, 2, \text{true})\) (line 9), since the first trace element matches the first event of block₁; 2) when \(i = 2\), \((i₁, t₁, \text{flag})\) is set to \((3, 6, \text{true})\) (line 10) and breaks the loop at line 13, since the function finds a match of block₁. Notice that the rest of the iteration does not impact the matching of block₁.

If the function has not yet found a valid occurrence of block₁, in the remainder of the iteration, it checks whether the current trace element is part of an occurrence of block₂ (lines 17–23). If an occurrence of the first event of block₂ is detected, the variable \(i₂\) is set to 2 and variable \(i₂\) is set to the timestamp of current trace element (line 17). Otherwise, if variable \(i₂\) is already greater than 1, the algorithm calls function match to match the current trace element with block₂ \([i₂]\) (line 18).

The algorithm reports a violation (lines 19–23) when block₂ is completely matched. More specifically, if an invalid occurrence of block₂ has not yet been found (i.e., \(l_w\) is still 0), it means that the pattern is violated because the occurrence of the second block is not matched by an occurrence of the first block; hence, an NSOR violation type is reported, together with the location of the current trace element (indicating the occurrence of block₂ just matched) and the null value indicating there is no corresponding occurrence of block₁ (line 20). Otherwise, a WTC violation type is reported, with the location of the current trace element (indicating the occurrence of block₂ just matched), and the location \(l_w\) of the corresponding invalid occurrence of block₁ (line 21). The diagnostic information is then added to variable result (line 22) and the tuples \((i₁, t₁)\) and \((i₂, t₂)\) are reset (line 23). After analyzing the entire sub-trace, the function ends by returning variable result (line 24).

5 TOOL SUPPORT

We have implemented our model-driven procedure for trace diagnostics in the tool TemPsy-Report, publicly available [12, 13]. This tool extends the implementation of our tool TemPsy-Check [18] and is also based on Eclipse OCL [20]. The tool works as follows: given a trace and a set of TemPsy properties (previously determined, by TemPsy-Check, as violated on the input trace), TemPsy-Report builds OCL queries following the template shown in Figure 1, depending on the type of the scope and pattern used in each TemPsy property. We have implemented all the OCL functions (defined on class Trace) for TemPsy scopes and patterns as described in the previous subsection, to collect diagnostic information. The evaluation of the OCL queries is done through the evaluate function.
6 EVALUATION

As stated by requirement R2 in Section 1, our solution for model-driven trace diagnostics is expected to support very large traces, with millions of events, such that violations in the trace could be processed within seconds. To check the fulfillment of this requirement, we evaluated the scalability of TemPsy-REPORT by investigating the relation between the execution time and some structural properties of a trace, such as the length and the number of violations. More specifically, we consider the following research questions:

**RQ1** What is the relation between the execution time of TemPsy-REPORT and the length (i.e., number of logged events) of a trace?

**RQ2** What is the relation between the execution time of TemPsy-REPORT and the number of violations (with respect to a given TemPsy property) contained in a trace?

### 6.1 Benchmark and Settings

**Benchmarks.** The benchmark for evaluating TemPsy-REPORT is constituted by a set of TemPsy properties and by a set of traces.

As for the properties, we employed the ones already used in our previous work [17], extracted from the requirements specification documents of an eGovernment application developed by our partner. Since we focus on the scalability of the trace diagnostics procedure, we considered the 12 (out of 47) properties that use the **globally** scope: the semantics of this scope guarantees that a pattern is analyzed through the entire length of the trace. These properties are shown in sanitized form in Table 1; for confidentiality reasons, we only keep the structure of each property, in terms of scope + pattern, and denote events with uppercase letters.

Regarding the traces, as already done in [17], we used synthesized traces, to cover a large spectrum of lengths while having a great, controlled diversity in terms of occurrences of violations in the traces. More specifically, by using synthesized traces we were able to control in a systematic way the factors (i.e., trace length, number and type of violations) required to answer the research questions, while setting other factors (e.g., distance between events) randomly, to avoid any bias.

We synthesized these traces using a trace generator program, which we also implemented. This program takes in input a TemPsy property and configuration options, and generates traces that violate the input property. The generator avoids bias by distributing the (events leading to) violations evenly, assigning them random positions within the slots determined by the input parameters, such as the trace length and the number of violations. The position and the order of the events related to a violation are generated randomly by taking into account the temporal and timing constraints prescribed by the semantics of the pattern used in the input property. Positions in the trace that are not related to the property are filled with a dummy, irrelevant event. In the following we briefly describe the trace generation strategy for each pattern; next to each pattern name, we also indicate the corresponding properties from Table 1.

**Univervality** (P1). There is only one possible violation type for this type of pattern: NSOC. The generator first randomly generates the violation positions and then inserts in them a dummy event; all the other positions will have the event indicated in the property.

**Existence** (P3, P4). The trace generation strategy depends on the bound indicated in the property, in terms of comparison operator and bound value m. If the bound is expressed as “at least m” (as in property P3), the violations will be of type NSOC; the number of their occurrences (i.e.,
the number of occurrences of the event indicated in the property) is set to the minimum between \(m - 1\) and the value of the input parameter "number of violations". If the bound is expressed as "at most \(m\)" (as in property P4), the violations will be of type UNOC. The number of the occurrences of the event indicated in the property is set to the maximum between \(m + 1\) and the value of the input parameter "number of violations". In both cases, the position of these events is set randomly; all other positions in the trace are filled with a dummy event. A similar process is followed for the other variants of the pattern, not used in our benchmark.

**Absence** (P2). This pattern can only be violated by triggering UNOC violations. The generator first randomly generates the violation positions and then inserts in them the event indicated in the property (e.g., \(b\) in the case of P2); all the other positions will be filled with a dummy event.

**Precedence** (P7–P10). For this type of properties, there may be more than one type of violations, depending on the structure of the pattern; to avoid any bias, the generator takes as input also the violation type to be generated and produces a set of traces containing only the required violation type. For example, all properties \(P7–P10\) can lead either to NSOR or to WTO violations. Given the number of violations \(n\) as parameter, the generator first divides the trace into \(n\) segments with the same length, and then randomly inserts a specific violation into a position within each segment, taking into account the distance constraints prescribed by the pattern. For example, to generate \(n\) NSOR violations, the generator produces \(n\) occurrences of the second block (e.g., the event chain "C, D" for property P10), without inserting the matching occurrences of the first block. To generate \(n\) WTO violations, the generator produces \(n\) pairs of the two blocks indicated in the property, making sure that their positions in the trace violate the constraint on the time distance between them. The strategy for choosing the value of the distance between the two blocks depends on the type of the comparison operator used in the distance constraint. If the bound is expressed as "at most \(m\)" (as in property P7), the distance is randomly generated using a uniform distribution on the range \([m + 1, m + 1.1]\); if the bound is expressed as "at least \(m\)" (as in properties P8 and P10), the distance is randomly generated using a uniform distribution on the range \([1, m - 1]\); if the bound is expressed as "exactly \(m\)" (as in property P9), the distance is randomly generated using a uniform distribution on the range \([1, \ldots, m - 1, m + 1, \ldots, m + 1.1]\). A similar strategy is used to generate violations of type WTC and WTC.

**Response** (P5–P6, P11–P12). The trace generation for this pattern is similar to the one described above for the precedence pattern, taking into account that response is the dual of precedence. Also in this case, the generator takes as input the violation type to be generated and produces a set of traces that contains only the required violation type. For example, properties P5, P6, and P11 can lead either to NSOR or to WTO violations, while P12 can lead only to an NSOR violation.

**Settings.** The results reported in this section have been measured (by invoking the `System.currentTimeMillis()` method of the standard Java library) on a desktop computer with a 3 GHz Intel Dual-Core i7 CPU and 16 GB of memory, running Eclipse DSL Tools v. 4.6.0M3 (Neon Milestone 3), JavaSE v. 1.8.0_25-b17, Java HotSpot (TM) 64-Bit Server VM v. 25.25-b02, mixed mode), and Eclipse OCL v. 6.0.1. All measurements reported correspond to the average value over 100 runs of the trace diagnostics procedure (on the same trace, for the same property).

### 6.2 Scalability Analysis

**Methodology.** To answer RQ1, for each property (and for each type of violation, NSOR and WTO, in the case of properties P5–P11), we generated ten traces with various lengths from 100K to 1M, with a 100K step increment; in the trace generator program we fixed the number of violations\(^2\) to 1000. To answer RQ2, for each property but P3\(^3\) (and for each type of violation, NSOR and WTO, in the case of properties P5–P11), we generated ten traces varying the number of violations\(^2\) from 1K to 10K, with a 1K step increment; in the trace generator program we fixed the length of the trace to 1M. In both cases, we ran TemPsy-Report to analyze, for each property, the ten corresponding traces.

**Results.** The plots in Figure 3 show the relation between the execution time of TemPsy-Report and the trace length (RQ1). The execution time of TemPsy-Report on a trace containing NSOR violations is denoted by adding a superscript \(^\dagger\) to the property name, while a superscript \(\ddagger\) is used to indicate the execution time on a trace with WTO violations. We split the plots into two parts (Figures 3a and 3b), to better highlight the execution time for trace diagnostics of the properties with a precedence pattern with a distance constraint of the form "at least \(n\) tu" (P8 and P10).

The answer to RQ1 is that the TemPsy-Report tool scales linearly with respect to the trace length; the execution time ranges from about 1.5 s to 8.2 s, depending on the pattern used in the property and the violation type in the trace.

The plots in Figure 4 show the relation between the execution time of TemPsy-Report and the number of violations (RQ2) contained in a trace. We use the same notation as above to distinguish between NSOR and WTO violations. Also in this case, the plots for the diagnostics of properties P8 and P10 are separated.

The answer to RQ2 is that the number of violations contained in a trace makes no tangible impact on the execution time of the TemPsy-Report tool, which stays approximately constant as the number of violations increases; the execution time ranges from about 3.8 s to 8.2 s depending on the pattern used in the property and the violation type in the trace.

We also gain additional insight by inspecting the plots in Figures 3 and 4: TemPsy-Report takes less time to collect diagnostic information for the properties using the universality, existence, and absence patterns (e.g., P1–P4) than for the properties using the precedence and response patterns. Also, Figures 3b and 4b show that the tool takes less time for analyzing traces with WTO violations than traces with NSOR violations, when considering properties with a precedence pattern with a distance constraint of the form "at least \(n\) tu" (P8 and P10). This is due to the implementation of the corresponding OCL function, which—in the worst case—has

\(^2\)In the case of the universality, absence, and existence patterns, the number of violations actually represents the number of occurrences of a specific event that lead to a violation.

\(^3\)Property P3 was not used to answer RQ2 because varying the number of violations does not make sense from a scalability analysis standpoint: the property is violated only when there is no occurrence or just one occurrence of event A.
to check the order of events blocks, the distance between them, and also the distance between the individual events within a block. Finally, the results also show that TemPsy-Report takes longer for collecting diagnostic information for an order pattern in which the blocks are event chains rather than single event (as in the case of P10 and P11); again, this is due to the implementation, which has to check for all events matching a single block and the corresponding time constraint between the individual events in the block.

**Threats to validity.** The main threat to validity to the results presented above is the intrinsic presence of errors in TemPsy-Report. We tried to compensate for this by thoroughly testing the tool with traces and properties for which the oracle (in terms of diagnostic information) was previously known. Another potential threat is the fact that we have performed trace checking on synthesized traces. Real execution traces might be different, in terms of violation types occurring in them, due to a different distribution of event occur-
references along the traces, both in terms of order and time distance. However, as explained at the beginning of this section, for the purpose of scalability analysis synthesized traces are better than real ones as they guarantee we have the data to perform our analysis by controlling certain factors and varying others randomly. Another threat is given by the use of Eclipse OCL; one could get different results by using another OCL tool, with lower performance. We chose Eclipse OCL for its scalability (see [36]).

Discussion. The evaluation results presented above show the feasibility of applying our model-driven approach for trace diagnostics, in realistic settings, of temporal properties expressed in TemPsy. The tool scales linearly with respect to the length of the input trace and keeps approximately constant performance with respect to the number of violations contained in the trace.

The performance of our TemPsy-Report tool, which can analyze very large traces (with one million events) in less than ten seconds, makes it a viable technology for adoption in contexts where MDE is the mainstream practice within the software development process.

7 RELATED WORK

Although run-time verification (and trace checking) has been a very active area of research in the last 15 years [4, 31], little or no attention has been paid to the problem of trace diagnostics. The work closest to our approach is the one by Ferrère et al. [23], which proposes an error diagnostics algorithm for trace diagnostics of Signal Temporal Logic (STL) formulae over a continuous signal; the algorithm computes temporal implicants, i.e., small sub-signals that are sufficient to imply violations. This algorithm has been recently integrated into the AMT 2.0 tool [35], which supports qualitative and quantitative analysis of hybrid continuous signals. The main difference with our approach is that AMT does not provide a model-driven approach, which is one of the main requirements (see Section 1) set by the context in which we have developed TemPsy-Report; furthermore, AMT focuses on signal-based, continuous time applications with STL specifications while TemPsy-Report considers event-based, discrete time traces and a restricted, pattern-based temporal specification language.

We have also surveyed the trace diagnostics support in the tools that were contestants of the “offline monitoring” track of the 2014 and 2015 international Competition on Software for Runtime Verification (CSRV 2014 [3] and CSRV 2015 [22]). Four of the tools (STePr, AgMon [29], LogFire [27], OptySim [8]) are not publicly available; another tool RiTHM-v2.0 [34], is available but does not work when executed by following the instructions specified in the README file on its GitHub page [39]; the remaining tools that we analyzed are: BREACH [10], MONPOLY [5], QEA [37], SOLOIST+ZOT [7], and RV-Monitor [33]. Table 2 summarizes our findings, indicating whether a tool produces a boolean output (column “boolean”), whether it takes into account all the (events leading to) violations in a trace (column “positions”), whether it provides detailed diagnostic information, such as the cause of a violation (column “cause”), and whether it is based on MDE technologies (column “MDE”). Although all tools yield a boolean result, only few provides additional information. BREACH, which can check digitized traces against Signal Temporal Logic (STL) specifications, provides a graphical visualization of the violations in the trace. MONPOLY prints out the last log entries read before finding a violation. QEA stops checking the trace after the first violation is found and prints the last log entry read before finding the violation. None of these tools provide additional diagnostic information to understand the cause of a violation. RV-Monitor does not report the position of the violations in a trace but it allows for manually writing a violation handler to print some user-(pre)defined diagnostic information. As seen in Table 2, only AMT and TemPsy-Report provide detailed diagnostic information to investigate faulty traces; however, only TemPsy-Report is a tool based on standard MDE technologies such as OCL.

Trace diagnostics is an activity related to falsification [9], e.g., finding counterexamples to a system specification, and to the analysis of counterexamples in the context of model checking [6, 38].

Finally, trace diagnostics can be seen as a particular case of query on a trace of events, as typically done in Complex Event Processing (CEP [32]) settings; strategies for the synergetic integration of RV and CEP are discussed by Hallé [26].

8 CONCLUSIONS

In this paper we described a model-driven approach for the diagnostics of the verdicts yielded after checking a property over a trace. We implemented our approach in TemPsy-Report, a tool that provides detailed trace diagnostic information for pattern-based temporal properties expressed in the TemPsy language; this information can be displayed in an interactive visualization tool.

The evaluation of TemPsy-Report shows that it can collect diagnostic information from large traces (with one million events) in less than ten seconds. TemPsy-Report scales linearly with respect to the length of the trace and is not affected by the the number of violations contained in the trace.

As part of future work, we plan to conduct a user study to assess how TemPsy-Report can support developers while performing fault localization, by providing useful diagnostic information.

ACKNOWLEDGMENTS

This work has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 694277), from the Luxembourg National Research Fund (FNR) (grant No FNR/P10/03), and from the University of Luxembourg (grant “MOVIDA”).

Table 2: Comparison of trace diagnostics features among trace checking/run-time verification tools

<table>
<thead>
<tr>
<th>Tool</th>
<th>boolean</th>
<th>positions</th>
<th>cause</th>
<th>MDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BREACH [10]</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MONPOLY [5]</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>QEA [37]</td>
<td>+</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SOLOIST+ZOT [7]</td>
<td>+</td>
<td>-</td>
<td>+/-</td>
<td>-</td>
</tr>
<tr>
<td>RV-MONITOR [33]</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>AMT [35]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>TemPsy-Report</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Model-Driven Trace Diagnostics for Pattern-based Temporal Specifications

REFERENCES


