Let’s Look at the Logs: Low-Impact Runtime Verification*

ALEX GROCE1, KLAUS HAVELUND2, MARGARET SMITH2, HOWARD BARRINGER3

1 School of Electrical Engineering and Computer Science
Oregon State University
Corvallis, USA

2 Jet Propulsion Laboratory
California Institute of Technology
Pasadena, USA

3 School of Computer Science
University of Manchester
Manchester, UK

Email:
alex@eecs.oregonstate.edu
{klaus.havelund,margaret.h.smith}@jpl.nasa.gov
howard.barringer@manchester.ac.uk

Runtime verification as a field faces several challenges. One key challenge is how to keep the overheads associated with its application low. This is especially important in real-time critical embedded applications, where memory and CPU resources are limited. Another challenge is that of devising expressive and yet user-friendly specification languages that can attract software engineers. In this paper, we show that for many systems, in-place logging provides a satisfactory basis for post-mortem “runtime” verification of logs, where the overhead is already included in system design. While this approach prevents on-line reaction to detected errors, possible with traditional runtime verification, it provides a powerful tool for test automation and debugging — in our case, analysis of spacecraft telemetry by ground operations teams at NASA’s Jet Propulsion Laboratory (JPL). The second challenge is addressed in the presented work through a temporal pattern language, designed in collaboration with JPL test engineers. The pattern language allows for descriptions of relationships between data-rich events (records) common in logs, and is translated into a form of automata supporting data parameterized states. The automaton language is inspired by the rule-based language of the RULER runtime verification system. We present a case study illustrating the use of our LOGS cope tool by software test engineers for the 2011 Mars Science Laboratory mission.

Received 08 Month 2009; revised 08 Month 2009

1. INTRODUCTION

Runtime verification encompasses the discipline of checking execution traces against formal specifications. Two of the key problems are: (1) how to obtain execution traces with minimal impact on observed software, and without over-burdening the user with the need to implement instrumentation and (2) how to design a specification language that is expressive yet user-friendly. This paper reports on an experiment addressing both these issues, a combination that we call “low-impact” runtime verification. In summary, the paper argues that analyzing log information removes the instrumentation burden in the sense that logs usually are already produced by most critical software systems. Consequently, also no additional impact on the running software is imposed beyond the already implemented logging. A log is a recorded sequence of events. We argue that systematizing logging can be of additional benefit. Second, we introduce a textual temporal logic inspired pattern language over events for specifying properties of logs. This language supports data parameterization, essential for monitoring logs (events

*Part of the research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.
typically carry data). Patterns are translated into data parameterized automata forming an interesting and useful subset of the textual RULER language [8, 10, 9]. Users can mix patterns with more expressive automata in a specification. Parameterized automata are visualized using the GraphViz tool [19]. Our system additionally offers preliminary support for automatically constructing (learning) specifications from example logs. The log analysis framework, LOGSCOPe, is developed to support engineers testing the flight software for NASA’s next Mars rover mission, the Mars Science Laboratory (MSL) [1], developed at the Jet Propulsion Laboratory (JPL). In this sense the work represents an instance of the often sought, but rarely found, marriage between theory and practice in formal methods.

The MSL mission’s goal is to put the so-far largest rover (the size of a compact car) on Mars for continued exploration of the red planet. It is scheduled to launch in 2011. A flight software team peaking at approximately 30 programmers develops the software for controlling the Rover Compute Element (RCE), which controls all stages of the integrated spacecraft. A testing team of approximately 10 engineers (the Flight software Internal Test, or FIT, team) is responsible for functional testing of the flight software. LOGSCOPe was developed to support this team, and is the result of their requirements as well as our research ideas. The MSL flight software produces rich log information, which is stored in SQL databases (one database per log). A log is in essence a sequence of time-stamped events, where an event can be one of several forms, corresponding to input to and output from the system, as well as internal state transitions and state readings. Each event is in essence a mapping from field names to values (a record). Such logs are usually far too large to be analyzed effectively by humans. Traditionally, these MSL logs have been analyzed by writing scripts, with properties coded up in PYTHON. Such scripts are time consuming to produce and result in difficult-to-read “specifications” that hinder communication, maintenance, and specification-sharing and reuse. Runtime verification of logs using formal specifications is a solution that has proven to be very effective for introducing formal methods into a high-profile NASA flight project.

1.1. Contributions and Related Work

The contribution of this work is a data-parameterized temporal pattern language, with a translation into parameterized automata (also directly usable in specification). A second contribution is our method for injecting this technology into a NASA flight mission. The LOGSCOPe system has specifically been influenced by the RULER system [8, 10, 9]. In particular, the automaton language conceptually forms a subset of RULER. LOGSCOPe can be seen as an adaptation of RULER to the specific needs of MSL; adding temporal logic, enhancing usability, and implementing it in PYTHON. LOGSCOPe has also been influenced by the state machine-based systems RCAT [27] and RMOR [20].

A substantial amount of work has been done on runtime verification [2] within the last decade, e.g. [22, 18, 21, 7, 4, 13, 17], much of which has guided the design of RULER as well as LOGSCOPe. Chen and Roşu [13] address the issue of designing the “right” logic by implementing a range of traditional logics, such as state machines, temporal logic, regular expressions, and grammars, as different plugins in the MOP system. A property must be stated in one of the different logics, a mixture is not possible. The LOGSCOPe logic on the other hand attempts to suggest the “right” logic for the specific purpose in hand. Data handling is handled uniformly in MOP by separating it out from the logic plugins, which are all propositional. This makes designing a new logic very easy since only the propositional case needs to be defined. However, this is attained at the cost of limited quantified specifications.

Software testing researchers [5, 12] have also explored formal specifications for checking logs. Chang and Ren [12] provide a pattern language just based on regular expressions. The properties are formulated over actual textual logs, in contrast to LOGSCOPe properties, which are formulated over an abstraction of the log: a sequence of records, each of which is a mapping from field names to values. This makes their approach less adaptable, less user-friendly, and more fragile to changes in the log format. Chang and Ren provide interesting empirical data on application to a large telecommunications system (though it is not clear if patterns are written by the testers or the research team).

Andrews and Zhang [5] share our focus on lightweight test oracles. They offer an elegant parameterized state machine framework similar to LOGSCOPe parameterized state machines. Their state machines yield errors in case events occur for which there is no outgoing transition. Although we support this concept, we also support state machines where one can wait in a state until a certain event occurs. This has turned out to be useful in practice for large log files and occasionally leads to less verbose specifications in the case studies we have looked at. Their state machines only support success states (final states) whereas we also support hot states. Their state machines are compiled into PROLOG. They do not provide a temporal logic, a concept that we found important for the engineers.

The PSL specification language [29] combines temporal logic and regular expressions, and in this way has commonalities with the LOGSCOPe language. However, PSL is designed for hardware verification and therefore does not support data parameterization to the degree that LOGSCOPe does. The focus on hardware verification also means that the next-operator becomes important, having as consequence that all the usual temporal operators must be introduced. The LOGSCOPe temporal logic language is characterized by integrating temporal logic and a limited form of regular expressions in a very simple manner (there are for example no explicit temporal operators). The LOGSCOPe specification language is much simpler than PSL. Engineers were able to write properties in less than one hour. Finally, LOGSCOPe offers parameterized state
mateescu and Thivolle [25] introduce a model checking language (MCL), which augments the modal μ-calculus with high-level operators aimed at improving expressiveness and conciseness of formulas. The main MCL ingredients are data parameterized fixed points, action patterns extracting data values from events, modalities on transition sequences described using extended regular expressions and programming language constructs. The attempt in [25] is similar to the work described in [7] on extending the μ-calculus with data and a sequencing (chop) operator for runtime monitoring. However, although very expressive we do not believe that formulas in MCL are necessarily easy to write for engineers. Our framework is simpler and directly applicable to formulate the kinds of properties needed for our case studies.

Cohen et al. [14] introduce the PDL specification language for stating properties about programs, to be verified using symbolic execution. The specification language has some similarities with our temporal pattern language in providing a form of sequencing operator and allowing for negation of events. However, PDL is limited in the alternation between positive and negative events allowed. More importantly, events can only be program labels. There is no support for data parameterization.

Dillon et al. introduce in [15] the Graphical Interval Logic GIL for specifying the behaviour of concurrent systems. The logic is quite powerful and addresses the the problem of usability of propositional temporal logic. However, the GIL logic is considerably more complex than the logic presented here, making fast learning by test engineers less likely. The LOGSCOPE logic is textual as opposed to GIL, which is a graphical notation. It has been our experience that test engineers work faster with a good textual notation rather than a graphical notation. It also makes it easier to auto-generate specifications. Finally, GIL does support some form of data quantification. However, there are no algorithms provided for monitoring such unless the type quantified over is finite and known [16].

The work in [3] describes a set of specification patterns often encountered in work on verification. Each pattern focuses on one specific property format, such as for example the response pattern (P should always be followed by Q). Our experience is that our general logic is simple enough for engineers to work with, and that special patterns are not necessarily needed. This is in fact the strength of the proposed logic in our belief. The specification patterns in [3] do not deal with data, and it would be a likely be a research topic to extend the patterns with such.

1.2. Outline of Paper

Section 2 describes the process that lead to the design of the LOGSCOPE specification language and system. Section 3 describes how log files are constructed for the MSL flight software, forming the foundation of a through-going example. Section 4 introduces the temporal logic pattern language, while Section 5 introduces the automaton language, and illustrates how patterns are translated to automata. Section 6 briefly discusses the problem of detecting typing inconsistencies between a log file and a specification. Section 7 briefly describes the learning feature. Section 8 describes the results of the MSL case study in more detail. Section 9 finally concludes the paper and suggests future work. Appendix A.1 contains the complete grammar for the LOGSCOPE specification language.

2. TOOL DEVELOPMENT PROCESS

The LOGSCOPE tool and pattern language grew out of the needs of engineers. In fact, the earliest effort to “mockup” a specification came from an MSL test engineer, not the runtime verification research team. The “specification” was a comment in a PYTHON script describing the purpose of the script as follows.

```
look:DRILL_DMP\n  evr(CMD_DISPATCH,positive)\n  evr(CMD_COMPLETED_SUCCESS,positive)\n  evr(CMD_COMPLETED_FAILURE,negative)\n  chan(id:CMD-0004,positive,contains opcode of last immediate command)\n  chan(id:CMD-0007,positive)\n  chan(id:CMD-0001,negative)\n  chan(id:CMD-0009,negative)\n  prod(name:DrillAll,1,*)
```

The script represents the following property to be checked: when an issued DRILL_DMP command is observed in the log, then the events with positive as an argument should follow in the given order, and the events with negative as an argument should not occur. For example, the following evr (event report) events should follow: a report of the dispatch (CMD_DISPATCH) of the command; a report of the success (CMD_COMPLETED_SUCCESS) of the command; and there should not be an evr reporting failure (CMD_COMPLETED_FAILURE) of the command. Then follows some requirements on samplings of the state (channel events). For example, there should be a sampling of state variable CMD-0004 that, informally stated, contains opcode of last immediate command. Finally the flight software should downlink1 a DrillAll product (a prod event) to ground informing about the status of the drill.

The core elements of our current pattern language — the events, all originate in the telemetry system of the Mars Science Laboratory flight and ground software. These events correspond to the generalized view of a “system” shown in Figure 1. An operations team on the ground issues commands to the spacecraft, e.g. DRILL_DMP in the sample specification. The spacecraft software responds

---

1“To downlink” is NASA terminology for sending information from spacecraft to ground.
to commands by causing transitions. Transitions in turn cause the local state to be updated. State changes can be observed either in terms of “random” state samples that provide a snapshot of current spacecraft state, or more precisely by observing the actual state changes when they occur. Finally, data products are the “outputs” of the spacecraft downlinked to the ground, including engineering telemetry, images, and science instrument data. These different event kinds are referred to in the logs (see Figure 2) as respectively: COMMAND, EVR (EVent Report = transition), CHANNEL (state sampling), CHANGE (state change) and PRODUCT. Although these MSL concepts are very general (should cover any system) LOGSCOPE can be adjusted to work with other forms of events.

To a large extent, the development of the pattern language that followed was a process of taking this early draft and preserving its essential elements while expanding its expressive power and readability, guided by usage, engineer requests, and language design considerations. For example, we preserved the structure of properties, each having the form: \( E \Rightarrow C \), where \( E \) is an event and \( C \) is a consequence, for example a sequence of events. However, more complex specifications were required. In the original mockup, the positive events in \( C \) were required to occur in order (allowing all non-negative events in between). The first implementation prototype introduced a way to distinguish ordered lists from unordered sets of events, which proved critical to real specifications.

The original test infrastructure, as suggested by the FIT team, provided an event registry, which was a PYTHON library with methods allowing test scripts to create, alter, and delete online monitors for telemetry events. In principle, the registry provided a set of tools sufficient for analyzing tests at runtime, and a specification and debugging mechanism well-integrated with established testing procedures. In practice, the registry approach was seldom adopted by engineers (and proved frustrating when used) for two primary reasons:

(i) **Ordering and timing of events**: Events on the spacecraft are not downlinked to the ground system in chronological order. Different channels, data products, and EVRs have different priorities, and the “Earth receive time” ordering of two events will often be the opposite of their event times. Test engineers were forced to either introduce lengthy delays after each test step or build very complicated logic to disentangle events arriving out-of-order and “recreate” a linear chronology of events (which is sometimes only possible after test termination). A strategy of pausing long enough to allow telemetry to arrive proved brittle, as timeouts fluctuated with each software release.

(ii) **Confusion of test execution and test evaluation**: A motivation for the event registry was that it seamlessly integrated with the idea of tests as scripts with logic, looping, and other programming language features. Unfortunately, test scripts that were clearly readable when their task was limited to commanding the spacecraft became difficult to follow when test execution was interfaced with test evaluation.

Neither of these problems is particular to MSL. The first problem is generic to any distributed system in which constructing an event timeline is non-trivial (before queues clear), and the second problem is a general observation about test case readability. Test engineers reacted to these problems by making little use of the registry, despite having requested its features. How, then, were test engineers evaluating tests? In many cases, they were hand-scripting a post-test analysis of all telemetry (ad-hoc construction of very limited logs). A few checks were routinely performed online, but more complex analysis was often delayed until the test was complete, and chronological confusion could be partially avoided by counting expected responses.

Team management (also active in writing tests) recognized that building ad-hoc test-specific “log analysis” systems was not an effective use of mission time, and that scripting logic only partially addressed both problems with the registry. The DRILL_DMP mockup emerged after a series of team meetings and discussions, informed by our ideas of log analysis. MSL flight software management suggested that a unified logging scheme would be useful to developers outside the test team. After an initial trial run in which the research team replicated the hand-scripted results of random command regressions, the FIT team began use of prototype versions of the logging system, replacing hand-scripted test evaluation code, as described in more detail in Section 8. The research team suggested the use of square and curly braces to indicate lists and sets of events, respectively, and the use of bindings for data values. Other features, including indexed data fields (e.g. bit-vectors), SQL query filtering of events, scoping of specifications (Section 4.6), and concrete learning (Section 7.1), were requested by the test team. A specification language emerged, starting with a syntax and
The first step in analyzing the behaviour of the MSL flight system is to produce a log — an ordered sequence of events.

3. LOGS

The first step in analyzing the behaviour of the MSL flight system is to produce a log — an ordered sequence of events.

Our log is based on a database view of commands and spacecraft telemetry maintained by JPL ground software. The ground software will be used by mission operations to communicate with the spacecraft and rover, and is now used both in MSL hardware testbeds and software simulation. The ground system stores events as entries in an SQL database (an approach also considered in runtime verification literature [23]). Different kinds of events in the database provide different fields (e.g., EVRs have a message but no value, while channels have a value but no message). Rather than forcing a single representation on these event types, we allow different events in the unified log to have varying fields, but ensure that each event has (1) a type and (2) a timestamp. The type is used in specifications as shown below, and the timestamp allows us to order the event sequence.

Figure 2 shows a simplified part of a log annotated by our tool after testing of the power sub-system. The example includes COMMAND, EVR, and CHANNEL events. The fields of the events are shown within curly brackets, except for the object type, which appears before the bracket. In addition to fields present in the original database, our system annotates events with derived fields that ease readability and specification. Field names from the MSL database begin in lowercase, while derived fields (e.g., Dispatch) begin in uppercase. The Time field, used to order events, is always derived. Events that take place on the spacecraft include a spacecraft event time (omitted here) that establishes a canonical order. However, command events originate from the ground, and include only a transmission time. We establish a uniform chronology by extracting the time a command is dispatched on the spacecraft from the EVR events — the Time field of COMMAND 7308 is extracted from the message in EVR 7309. This leads to a feedback between logging and telemetry design on the spacecraft. Use of our tool by test engineers has increased the MSL software team’s awareness of ambiguities in timing of events originating on the spacecraft, suggesting that improved synchronization is needed between modules responsible for different types of telemetry.

From the point of view of LOGSCOPE, a log is simply an ordered list of structures (implemented as a sequence of dictionaries in PYTHON). All MSL-specific aspects of the log are encapsulated in the logging tool, making the later tools in the chain easily adaptable to any system producing such a log. We expect this to be useful for other JPL missions, but it should also be applicable to operating systems, web servers, and any other systems that produce event-based logs. More generally, we note that even when a system produces logs in a purely textual format, an intermediate layer that translates this into a unified database view may be critical to maintenance of the log analysis framework: adapting hand-built regular expressions in each test to changes in logging output is time-consuming and error-prone. Constructing specifications that largely rely on derived fields is simpler for testers and encapsulates all parsing and interpretation of log events in one maintainable tool, the logging tool.

---

**FIGURE 2.** A simplified log
4. THE PATTERN LANGUAGE

Monitors are written in the LOGSPEDE specification language. A specification consists of one or more
specification units, each of which is either a temporal logic pattern, or a parameterized automaton:

\[
\langle \text{specification} \rangle \rightarrow \langle \text{monitorspec} \rangle^+
\]

\[
\langle \text{monitorspec} \rangle \rightarrow \langle \text{pattern} \rangle | \langle \text{automaton} \rangle
\]

In this section we focus on temporal patterns, which are automatically translated into parameterized automata.

4.1. Simple Response Patterns

Consider as an example the following informally stated property:

\[ R_1: \text{"Whenever a flight software command is issued, then eventually an EVR should indicate success of that command".} \]

Before we can formalize this property, it needs to be refined to refer to the specific fields of events. The following is such a refinement:

\[ R'_1: \text{"Whenever a COMMAND is issued with the Type field having the value "FSW", the Stem field (command name) having some unknown value } x, \text{ and the Number field having some unknown value } y, \text{ then eventually an EVR should occur, with the field Success mapped to } x \text{ and the Number field mapped to } y".} \]

The Number field is a sequence number indicating the order in which commands are received by the spacecraft for dispatch. Subsequent commands should have increasing numbers. Events related to command execution have the number of the command. In our language, this property reads:

\[
\text{pattern } \text{P1:}
\]

\[
\begin{align*}
\text{COMMAND}\{\text{Type:} \text{"FSW"}, \text{Stem:} x, \text{Number:} y\} & \Rightarrow \\
\text{EVR}\{\text{Success:} x, \text{Number:} y\}
\end{align*}
\]

This pattern (pattern is a keyword) has the name P1 and states that if a command is observed in the log at a position \( i \), with the Type field having the exact string value \"FSW\", the Stem field having some value \( x \), and the Number field having some value \( y \); then later in that log, at a position \( j > i \), an EVR should occur with a Success field having \( x \) as value and a Number field having \( y \) as value. Informally, we can explain the semantics of this formula in terms of the following formula in a Linear Temporal Logic (LTL) [26] with quantification over data variables:

\[
\forall x, y \bullet \\
\Box (\text{COMMAND}\{\text{Type:} \text{"FSW"}, \text{Stem:} x, \text{Number:} y\} \Rightarrow \\
\Diamond (\text{EVR}\{\text{Success:} x, \text{Number:} y\}))
\]

The temporal operators \( \Box \) and \( \Diamond \), as well as the universal quantification \( \forall \), are implicit in LOGSPEDE's pattern notation. The pattern has the form:

\[
\langle \text{pattern} \rangle \rightarrow \\
\langle \text{NAME} \rangle \ast : \ast (\langle \text{event} \rangle \ast : \ast (\langle \text{consequence} \rangle)
\]

\[
\langle \text{consequence} \rangle \rightarrow \\
\langle \text{event} \rangle \\
| \ast \langle \text{event} \rangle \\
| \ast \langle (\langle \text{consequenceclist} \rangle) \ast \rangle \\
| \ast \langle (\langle \text{consequenceclist} \rangle) \ast \rangle
\]

\[
\langle \text{consequenceclist} \rangle \rightarrow \\
\langle \text{consequence} \rangle | \ast, \ast (\langle \text{consequence} \rangle)
\]

This example shows the simplest case, where the consequence is an event. The other alternatives will be explained in the following subsections. The event triggering the pattern is a command event here, but can be any kind of event. Each event is constrained (between \{ ... \} brackets) by zero or more constraints, each consisting of a field name (without quotes), and a range specification. We saw two forms of range specifications: the string \"FSW\" for the field Type and the names \( x \) and \( y \) for the other fields. A string constant represents a concrete constraint: the field in the event has to match this value exactly (by PYTHON equality \==). One can also provide an integer as such a concrete range constraint. Ranges can also be multi-valued, such as integer intervals, as will be shown below.

A unnamed value \( x \) and \( y \) in this case occurring as a range indicates, at specification time, an unknown value. The first occurrence of such a name in a pattern is binding: it will be bound to the value of the corresponding field in the matching event. Any subsequent occurrences of this name in the pattern are now constraining: for a match to occur, the corresponding fields now have to have the values these names were bound to by the binding event. The log in Figure 2 satisfies this specification for the first command, as the command is matched by a success. The property fails for the second command.

4.2. Negation of Events

A consequence can also be the negation ('!') of an event. Suppose we want to state the following property:

\[ R_2: \text{"Whenever a COMMAND is issued with the Type field having the value "FSW", the Stem field having some value } x, \text{ and the Number field having some value } y, \text{ then there should thereafter not occur an EVR, with the field Failure mapped to } x \text{ and the Number field mapped to } y".} \]

We express this by the property (which the log from Section 3 satisfies):

\[
\forall x, y \bullet \\
\Box (\text{COMMAND}\{\text{Type:} \text{"FSW"}, \text{Stem:} x, \text{Number:} y\} \Rightarrow \\
\Diamond (\text{EVR}\{\text{Success:} x, \text{Number:} y\}))
\]
4.3. Composite Consequences

We have seen that the consequence of a pattern can be an event, as in pattern P1, or the negation of an event, as in pattern P2. There are two more forms: ordered and unordered sequences of consequences (a recursive definition). The square brackets [...] indicate that the consequences should occur in exact order, while the curly brackets {...} indicate that they may occur in any order. The symbols are chosen for their frequent use in literature to represent ordered lists and unordered sets. As an example, consider the following requirement:

R3: “Whenever a flight software command is issued, there should follow a dispatch of that command, then exactly one success of that command. Before the dispatch there should be no dispatch failure, and in between the dispatch and the success there should not be a failure”.

Requirement R3 is in Figure 3 illustrated as a timeline, much like in Timeedit [28], showing events that could/should occur, and constraints on events that should not occur. Note, however, that Timeedit does not support data parameterization, only temporal ordering of atomic events. This property can be stated formally as follows, in a form that very closely reflects the time line.

The consequence consists of a sequence (in square brackets [...] of (sub) consequences, in this case events and negations of events. The ordering means that as a response to the command the dispatch should occur before the success, and the negations represent what should not happen in between the non-negated events.

The following first order LTL formula represents this property. It is suggestive as to the extra complication that LTL’s until operator (‘U’) introduces, and why the pattern language might be easier to use for engineers:

\[
\forall x, y. \\
\square(\text{COMMAND\{Type:"FSW",Stem:x,Number:y\} \Rightarrow \neg \text{EVR\{DispatchFailure:x,Number:y\}}} \\
\text{U} (\text{EVR\{Dispatch:x,Number:y\}} \\
\land (\neg \text{EVR\{Failure:x,Number:y\}} \text{ U} \\
\text{EVR\{Success:x,Number:y\}} \\
\land \square \neg \text{EVR\{Success:x,Number:y\}})))
\]

As an example of an un-ordered arrangement of events, consider the following (perhaps less meaningful) relaxation of the above stated property:

R4: “Whenever a flight software command is issued, there should follow a dispatch of that command and exactly one success of that command, in no particular required order. There should be no dispatch failure and no failure at any time”.

We formalize this as:

The curly brackets {...} indicate an unordered collection of consequences (a Boolean ‘∧’ effectively). This corresponds in this case to four parallel timelines, all
of which must be satisfied. The fact that they are unordered means that the non-negated events can occur in any order (but must occur eventually), and negations have to hold to the end of the log. Nested inside the unordered \{\ldots\} construct there is, however, an ordered sequence expressing that after a success there should not occur another success. As the grammar for consequences suggests, ordered and unordered collections of consequences can be mixed arbitrarily.

\subsection{4.4. Event Predicates}

Events can be constrained with predicates. Consider the requirement:

\begin{quote}
R5: “The success of a command with a number y should never be followed by the success of a command with an equal or lower number \( z \leq y \)”.
\end{quote}

The following pattern formalizes this requirement.

\begin{verbatim}
pattern P5 :
  EVR{Success:_,Number:y} =>
  !EVR{Success:_,Number:z} where \{ : z <= y : \}
\end{verbatim}

Note that this formula ignores which commands succeed (the underscore ‘_’). The constraining predicate is: \{ : z\leq y : \}, and is a general \texttt{PYTHON} expression enclosed by the symbols \{ : \ldots : \}. This expression can refer to all of \texttt{PYTHON}’s directly available features, in addition to a \texttt{LOGSCOPE}-specific pre-defined library of predicates. Should this not suffice, it is possible to import \texttt{PYTHON} libraries or directly define \texttt{PYTHON} predicates in the specification. The following specification defines a \texttt{PYTHON} predicate \texttt{within(t1,t2,max)} which checks that the two time points \( t1 \) and \( t2 \) are at most \( max \) time units apart.

\begin{verbatim}
{:
  def within(t1,t2,max):
    return (t2-t1) <= max
  :}
\end{verbatim}

\begin{verbatim}
pattern P6:
  COMMAND{Type:*FSW*,Stem:*PICT*} =>
  COMMAND{Status1:{0 : 1, 4 : x}},
  COMMAND{Status8:{2 : x}},
  PRODUCT{ImageSize:[1000,2000]}
\end{verbatim}

The requirement states that after a flight software picture command, there should be a channel reading where the \texttt{Status1} field denotes an integer, where bit number 0 is 1 (counted from the right), and where bit 4 is bound to \( x \) (0 or 1). Thereafter should follow another channel reading where the \texttt{Status8} field denotes an integer, where bit 2 has the value \( x \). Finally should follow a product, where the field \texttt{ImageSize} is in the interval 1000 to 2000. The form of indexing shown for the status fields (\texttt{Status1} and \texttt{Status8}) generally works on any data object that is indexable, such as bitvectors (integers), lists and strings (in both cases indexes must be integers, counting from the left), and dictionaries (maps, allowing for indexing also with strings).

\subsection{4.5. Event Actions}

Events can be associated with actions to be executed when the events match. As an example, consider the following specification, which is a weakening of property P1. It states that only the first command must be followed by a success \texttt{EVR}.

\begin{verbatim}
{:
  counter = 0
  def count():
    global counter
    counter = counter + 1
  def amongstfirst(limit):
    return counter < limit
  :}
\end{verbatim}

\begin{verbatim}
pattern P8:
  COMMAND{Stem:x}
  where amongstfirst(1) do count() =>
  EVR{Success:x}
  do {:
    print x + " succeeded" :
  }
\end{verbatim}
The specification first introduces a PYTHON code section: a global integer counter, a count() function, and an amongstfirst(limit) function. If a command still matches the where predicate (it is amongst the first 1 commands), the do construct is executed, causing the count() function to be called, counting up the global counter variable. Just as predicates can be referred to by name (without enclosing in { ... }), so can actions as in this example. The general format of an event is:

\[
\langle \text{event} \rangle \rightarrow
\]

\[
\begin{align*}
\text{(type)} & \{ \ast \text{ (constraints)} \} \ast \\
\text{[where (predicate)]} \\
\text{[do (code)]}
\end{align*}
\]

If the user-defined PYTHON code contains the definition of a method final() (without arguments), then this method will be called at the end of monitoring the properties. This function can for example compute and print statistics based on the state of the global variables introduced. Each specification file has its own PYTHON state. Hence, it is possible to declare a state local to a subset of properties.

4.6. Scopes

In some cases one may want to limit the scope in which a pattern is checked, by providing an additional scope-terminating event. Without such limitations a pattern holds from the point at which its trigger event is matched until the end of the log. As an example, one may want to check that a particular command results in a particular set of events to occur, and some other events not to occur, up to the next command being fired. Consider for example pattern P4 in Section 4.3. According to the semantics, whenever a flight software command is detected in the log, the consequence is checked on the rest of the log, to its end. That is, any required event such as the dispatch can occur anywhere in the rest of the log, and negative events, such as failures are checked on the remaining log. We might, however, limit these checks to be performed up to the next flight software command (satisfying \text{COMMAND}\{Type:"FSW*"\}). This is done by adding the scope delimiter \text{upto} \text{COMMAND}\{Type:"FSW*"\} as follows:

\[
\text{pattern P9 :} \\
\text{COMMAND}\{Type:"FSW*,Stem:x,Number:y\} \rightarrow
\]

\[
\begin{align*}
\text{EVR}\{\text{Dispatch}:x,\text{Number}:y\}, \\
\text{[EVR}\{\text{Success}:x,\text{Number}:y\}, \\
\text{!EVR}\{\text{Success}:x,\text{Number}:y\}\}
\end{align*}
\]

\[
\text{upto} \text{COMMAND}\{Type:"FSW*"\}
\]

This means that positive events such as the dispatch have to occur before the next flight software command, and negative events, such as failures, are only checked for (forbidden) up to the next flight software command. The syntax for patterns is now expanded to include an optional scope-specification:

\[
\langle \text{pattern} \rangle \rightarrow
\]

\[
\text{pattern} \langle \text{NAME} \rangle \ast ; \ast (\text{event}) \ast \rightarrow (\text{consequence})
\]

\[
\text{[upto} (\text{event})\]
\]

5. THE AUTOMATON LANGUAGE

\text{LOGScope} also allows testers to write properties as parameterized finite-word \forall-automata, a more expressive, but lower-level language than the pattern language. An \forall-automaton is characterized by allowing several states to be active at the same time, all of which have to lead to success. In traditional non-deterministic \exists-automata many states are allowed active at the same time, but only one state has to lead to success. Patterns are automatically translated to automata. The automaton language forms a subset of the powerful rule-based RULER specification language [8, 10, 9], with some modifications. The automaton language has been designed to be sufficiently expressive to support the actual encountered needs of the MSL testing engineers, and yet as simple as possible to minimize learning effort. RULER is a much more comprehensive specification language, resembling a functional stream programming language. The differences between \text{LOGScope} and RULER are explained at the end of this sub-section.

In \text{LOGScope} an automaton is expressed in terms of states and transitions between states triggered by events. Events are exactly as defined for patterns. Just as events can be parameterized with values, states can be parameterized, carrying values produced by incoming transitions. Automata, hand-written as well as translated from patterns, can be visualized with GraphViz [19]. Automata are stored in GraphViz’s dot format for diagrams. Automata visualization has proven useful for users of the pattern language when trying to ensure their patterns express what is intended — we believe that textual development with graphic visualization is a very effective method for experienced test engineers. We illustrate the automaton language by presenting the automata for patterns P3 and P4.

5.1. Automaton for Pattern P3

The automaton corresponding to pattern P3 is shown in Figure 4, and visualized in Figure 5. The automaton consists of four states: S1, S2, S3 and S4, where the first mentioned state is the initial state (can also be identified via an initial keyword, as there may be several initial states). There is one transition exiting S1: this transition is triggered by a flight software command, binding x and y, and entering state S2(x,y) with x and y now bound to the actual values in the matching event — an example of a state parameterized with data. The S1 state is an always-state, meaning that it is always active, waiting for any command observed. As such
5.1. Automaton for Pattern P3

The automaton A\textsubscript{P3} for property P3 is defined as follows:

\begin{verbatim}
automaton A_P3 {
    always S1 {
        COMMAND{Type:"FSW",Stem:x,Number:y} => S2(x,y)
    }
    hot state S2(x,y) {
        EVR{DispatchFailure:x,Number:y} => error
        EVR{Dispatch:x,Number:y} => S3(x,y)
    }
    state S3(x,y) {
        EVR{Failure:x,Number:y} => error
        EVR{Success:x,Number:y} => S4(x,y)
    }
}
\end{verbatim}

Several instances of $S2(x,y)$ can be active at any point in time, each binding different values to $x$ and $y$.

States $S2$ and $S3$ are so-called hot states, meaning that instances of these states should disappear before the end of the log is analyzed. If not it is regarded as an error. They are used to model (bounded) liveness properties: that some event must occur before the end of the log. It is also possible to declare states hot with a separate hot-state declaration, as in: ‘hot S2, S3’. In this case, $S2$ for example represents the property that a dispatch should occur (with no previous dispatch failure). The first exiting transition is matched by any EVR with a DispatchFailure field with a value equal to the parameter $x$ and a Number field with a value equal to the parameter $y$. Similarly for the second transition. There are two special states: the error state (indicating error) and the done state (indicating successful termination of a branch of the automaton). The state $S4$ is a normal state (not an always-state and not a hot-state), meaning that it is acceptable to finish monitoring in that state. Its purpose is to check that there no further successes after the first, for a particular command $x$ with number $y$.

Automaton visualization uses different symbols for the different states: a grey state represents an initial state, a white circle represents a normal state, a black circle represents an error state, and a downwards-pointed red pentagon (in case colors are available) represents a hot state. The ‘@’ symbol in state $S1$ indicates an always-state.

5.2. Automaton for Pattern P4

The pattern P4 is translated into the automaton in Figure 6, and visualized in Figure 7. This automaton expresses that after a command, we want to see a dispatch (state $S2$) and a success (state $S3$), and after that no further success (state $S6$), but in no particular order, and that there should not at any time be a dispatch failure (state $S4$) or a failure (state $S5$). The transition of state $S1$ enumerates several target states, all of which become active when the transition fires, and all of which must lead to success: no error states should be entered and at the end no hot states should be active. The states can

\begin{verbatim}
automaton A_P4 {
    always S1 {
        COMMAND{Type:"FSW",Stem:x,Number:y} => S2(x,y), S3(x,y), S4(x,y), S5(x,y)
    }
    hot state S2(x,y) {
        EVR{Dispatch:x,Number:y} => done
    }
    hot state S3(x,y) {
        EVR{Success:x,Number:y} => S6(x,y)
    }
    state S4(x,y) {
        EVR{DispatchFailure:x,Number:y} => error
    }
    state S5(x,y) {
        EVR{Failure:x,Number:y} => error
    }
    state S6(x,y) {
        EVR{Success:x,Number:y} => error
    }
}
\end{verbatim}
be said to “execute in parallel”. The visualization of this automaton uses an upward pointing blue triangle (similar to a ‘∧’) to describe a transition with multiple targets.

5.3. Other Language Concepts

The specification language offers a number of additional features. These include a notion of success states, a dual to hot states. Occasionally it can be more convenient to express a property using success states: at least one of these must be reached by the end of the log analysis. States can also be defined as “step” states, meaning that in case the next event does not trigger one of the transitions, monitoring along that path is terminated. This is not an error, unless it means that some success state is not reached. This concept is specifically used during learning.

The language supports both C-style multi-line comments ({/*...*/}) and PYTHON-style single-line trailing comments (#...). Furthermore, it is possible to ignore a pattern or automaton by prefixing the declaration with the keyword ignore. This is useful for switching specifications on and off, as an alternative to commenting them out.

As mentioned earlier, the automaton language is more expressive than the pattern language. Recall that an event pattern has the form: event => consequence. That is, the antecedent is a single event. It is therefore currently not possible to write a pattern of the form: [E1, E2, E3] => E4, with the intended meaning that if events E1, E2 and E3 occur, then E4 should occur, despite this property being easily expressed in the automaton language. The pattern language can, of course, be extended to also allow this form of property (work in-progress).

5.4. Relationship to RULER

As mentioned above, the LOGSCOPE automaton language forms a simplified subset of the RULER language needed for translating the pattern language. As such it represents an interesting simple but powerful language, focusing on data parameterization, for writing monitors. The RULER concepts that have been ignored are the following. Whereas LOGSCOPE supports state machines where a transition in one state is triggered by one event and leads to a conjunction of resulting states, RULER is indeed a rule-based system where a transition can be triggered by a conjunction of states, events, and negations of such; and where the result is a disjunction of conjunctions of states and negations of such (negating a state means removing it). Disjunction is in RULER used to represent alternative ways of achieving success. RULER in addition has a notion of super state, similar to that in statecharts/UML. In RULER, templates (what correspond to LOGSCOPE automata) can themselves be parameterized with data as well as with other templates, making RULER higher order. Furthermore, various combinators exist for composing templates (sequential composition, conditionals, parallel composition, etc.).

6. TYPE CHECKING

In addition to the types of events, we must consider the types of event fields. User specifications are statically type-checked to ensure consistencies within a specification; we omit the details of this very standard (and, in this case, fairly trivial) process. Unfortunately, static checking does not detect the most common and serious type errors: inconsistencies between a specification and a log. Because we provide a very flexible specification language, agnostic as to the deep structure of events, we must handle these dynamic type errors at run-time, when we can compare with a log. Dynamic type errors can be divided into two classes: field errors and range errors. In field errors, a field is mentioned in the specification but is not present in the log, often due to misspellings or (less commonly) undocumented changes to the log database structure by the MSL ground system or flight software. Range errors generally concern a case where the specification expects one type (i.e., a string, integer, or list) and the log provides a different type object for the field. Another form of range error is when a specification indexes an event field, but the index is out-of-bounds (negative or larger than the size of the object). Our (somewhat ad hoc) solution is to provide users with a list of all patterns and automata that never trigger (which may indicate an error in a field’s name, or simply an event that did not occur) and all range violations after checking a specification. In order to improve usability (and in the spirit of PYTHON) we also provide limited automatic conversions (e.g., of strings containing numbers to numbers) while still reporting the range error in question.

7. LEARNING

Writing specifications can be difficult and time consuming. The key problem consists of identifying what properties to check — even formulated in English prose. A working approach is to generate logs during one or more runs,
and then look at (eyeball) the logs while trying to extract correct-behaviour patterns. This process can be supported by tools in various ways. One extreme approach is to attempt to learn automatically a specification from one or more presumably "well behaved" logs, and then later turn this specification into a monitor to check subsequent logs as part of a regression test suite. If these differ, for example due to later code modifications, warnings highlighting the discrepancies are issued. This approach assumes that a user can judge whether an initial set of logs (those to be learned from) reflect correct program behaviour, and is, of course not a generally sound approach, especially if logs contain thousands of events. If the initial logs are the result of erroneous program behaviour, the learner will learn a wrong specification. However, as long as error reports are considered with care, this technique is effective for catching bugs and "characterizing" software that is in a constant state of flux. LOGSCE provides some preliminary functionality supporting learning of specifications from logs. Two forms of learning can be considered: concrete learning and abstract learning. Only concrete learning has been implemented so far, but we will briefly discuss both.

**7.1. Concrete Learning**

During concrete learning LOGSCE learns the exact logs it observes up to equivalence on a set of field names that is provided to the learner, either from a default set or a set (for each kind of event) provided by the user. Two events are considered equivalent if they are equal with respect to these fields. The learner module, when given several logs, will build an automaton which represents the set of all logs seen. Common prefixes of the logs will result in a single path through the automaton, which will eventually form a tree-shape (no loops). The leaf-states are success states (at least one success state must be reached before the end of the log), and the remaining states are step states (states that must be left in the next step, hence forcing complete conformance to the automaton, no extra events are allowed). It is possible to create a new automaton, learn from one or more logs, write back the automaton learned so far to persistent memory, and then later further refine via learning or use it for monitoring. Concrete learning results in typically very large automata. The objective is in this case similar to that of comparing logs with UNIX's diff command. Here one would possibly also attempt to remove irrelevant event fields, for example with UNIX's grep command. Using an automata-based approach, however, opens up the possibility of more advanced learning of logs, what we term abstract learning.

**7.2. Abstract Learning**

In abstract learning mode one would want, for example, to learn high level properties about consequences of individual commands, yielding typically small, human readable, specification units, an idea similar to the approach of Perracotta [30]. For example, for each command one can learn what subsequent events always occur between the command and the next command. This corresponds to introducing a scope for learning the consequences of a single command. One can learn the specific order in which subsequent events occur or ignore the order. One can learn minimal and maximal time periods in between events. We also hope to incorporate classic automata learning results [6, 11], though the challenge here is considerable: the language of input symbols is very large, it is not feasible to make precise language-inclusion queries as required by Angluin's algorithm, and we believe that the SAT problems for Biermann's approach might be too large to solve.

## 8. USAGE AND CASE STUDIES

### 8.1. Running LogScope

LOGScope is called from a PYTHON application as in the following script, which first creates a log (typically by reading in a log generated by a running program), then creates an instance of an Observer object (given the location of the specification(s), possibly in a list, as parameter), and then applies the monitor method to the log:

```python
import logsce

log = ... # create a log
observer = logsce.Observer("$MSL/spec")
oobserver.monitor(log)
```

Executing the above PYTHON script will generate a file with results as well as a set of dot-files visualizing the generated automata. The results of the monitoring can also be accessed from within the PYTHON script by a getResults() method, so that they can be processed, e.g. as part of a regression test harness.

Our original log in Figure 2 violates some of the properties presented in this paper. This is summarized by the system as follows:

```
<table>
<thead>
<tr>
<th>Specification</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>1 error</td>
</tr>
<tr>
<td>P2</td>
<td>0</td>
</tr>
<tr>
<td>P3</td>
<td>1 error</td>
</tr>
<tr>
<td>P4</td>
<td>3 errors</td>
</tr>
<tr>
<td>P5</td>
<td>0</td>
</tr>
<tr>
<td>P6</td>
<td>0</td>
</tr>
<tr>
<td>P7</td>
<td>0</td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
</tr>
<tr>
<td>P9</td>
<td>3 errors</td>
</tr>
<tr>
<td>A_P3</td>
<td>1 error</td>
</tr>
<tr>
<td>A_P4</td>
<td>3 errors</td>
</tr>
</tbody>
</table>
```

All these violations are caused by the dispatch failure (and subsequent lack of a dispatch and a success) of the last...
command issued (event number 9626). To simplify this presentation we will focus on property \texttt{A\_P4} presented in Section 5.2, and visualized in Figure 7. This property is similar to \texttt{P4}. \texttt{A\_P4} is violated 3 times: once for a safety property (a failure occurs) and twice for a (bounded) liveness property (neither a dispatch nor a success occurs). The error messages for the safety violation and the first liveness property (lack of dispatch) are listed below.

\begin{verbatim}
command issued (event number 9626). To simplify this presentation we will focus on property \texttt{A\_P4} presented in Section 5.2, and visualized in Figure 7. This property is similar to \texttt{P4}. \texttt{A\_P4} is violated 3 times: once for a safety property (a failure occurs) and twice for a (bounded) liveness property (neither a dispatch nor a success occurs). The error messages for the safety violation and the first liveness property (lack of dispatch) are listed below.

\begin{verbatim}
RESULTS FOR A\_P4:

*** violated: by event 9627 in state:

state S4(x,y) { EVR{DispatchFailure:x,Number:y} => error }
with bindings:
{'x' : 'RUN\_COMMAND', 'y' : '18'}

by transition 1:
EVR('DispatchFailure' : 'RUN\_COMMAND', 'Number' : '18') => error

--- error trace: ---

COMMAND 9627 {
  Args := ['set\_dev(1)', 'TRUE']
  Number := "18"
  Stem := "RUN\_COMMAND"
  Time := 51708372934400
  Type := "FSW"
}

EVR 9627 {
  name := "CMD\_DISPATCH\_VALIDATION\_FAILURE"
  level := "COMMAND"
  DispatchFailure := "RUN\_COMMAND"
  Time := 51708372934499
  message := "Validation failed for command RUN\_COMMAND: number=18."
}

*** violated: in hot end state:

state S2(x,y) {
  EVR{Dispatch:x,Number:y} => done }
with bindings:
{'x' : 'RUN\_COMMAND', 'y' : '18'}

--- error trace: ---

COMMAND 9626 {
  Args := ['set\_device(1)', 'TRUE']
  Number := "18"
  Stem := "RUN\_COMMAND"
  Time := 51708372934400
  Type := "FSW"
}
\end{verbatim}

The first error message explains the safety violation: that the dispatch failure event number 9627 triggers the error transition in the state \texttt{S4}, which is listed, including the values that are bound to the state parameters \( x \) and \( y \). The error trace only contains the events that have made the automaton move: one dispatch failure event for the \texttt{RUN\_COMMAND} command event, and the fatal dispatch failure event for that command. Since events are numbered it is easy to locate these events in the real log, which is also stored with event numbers.

The second error message explains the first liveness violation: a dispatch does not occur. The \texttt{hot} state \texttt{S2} is where the monitor rests at the end of the log, indicating a violation. The only event that moved the monitor with respect to this particular violation is the initial \texttt{RUN\_COMMAND} command, which is the only event listed in the error trace.

\subsection{8.2. Application in MSL Testing}

As described, our tool and language development were guided generally by the needs of the MSL test engineers. Our first trial run, initiated at the request of MSL software management, checked the core behaviour of the command dispatch and success protocol shown above as pattern \texttt{P3}, for commands issued in randomly generated groups of 400 (an automated regression effort pre-dating our tool). Using LogCOPE enabled us to experiment with the specification more than the test engineer using a hand-scripted analysis, leading to the discovery of a previously uncaptured error, namely, duplicated success EVRs. A more extensive application, initiated by the test engineer responsible, was the automatic generation of a specification from tests of the power module. The test engineer replaced previously unreliable on-the-fly queries to the ground software with code to generate a specification for each tested behaviour. Enabling this automatic generation was a primary motivation for several language features, including pattern scopes. Again, log analysis revealed several faulty behaviours in the power and command modules, and revealed subtle issues with the timing of channel telemetry. The sample log in Figure 2 is taken from a power test (the full log contains over 11,000 events, including 107 flight software commands). The test engineers responsible for the file verification system (FVS, used when ground sends files via an uplink to the spacecraft) and the PYRO system (used to fire pyros) also replaced hand-scripted test code with LogCOPE specifications. In the case of the PYRO module, the test engineer needed the capability to learn a canonical log and compare other logs against the known-correct result, a primary motivation for the concrete learning capabilities of the tool.

MSL test team management has proposed that all scripted test evaluation should be converted to use LogCOPE, in order to simplify reaction to future changes in flight software, improve test maintainability, and enable automatic
overnight regressions of flight software behaviour. We believe that developing the tool in collaboration with the test engineers (and the larger MSL flight software team) has maximized the tool’s adoption and utility, and can serve as a model for the introduction of formal specification methods in other software efforts. More importantly, it seems clear that the tool has improved the test team’s productivity, and will result in better-tested flight software.

9. CONCLUSIONS AND FUTURE WORK

We have in this paper presented a log analysis specification language offering two forms of specifications: temporal logic patterns and ∀-automata, both of which can be parameterized with data, primitive and composite (lists, maps). The automaton language forms conceptually a subset of the RULER language, limited to the concepts needed by MSL. In this sense, the automaton language presented here suggests a useful, and yet simple subset of RULER. The automaton language is more expressive than the pattern language and can be used in cases where this extra expressiveness is required. Patterns are automatically translated to automata. To the best of our knowledge the translation from parameterized patterns to parameterized automata is a novel approach. A point of particular interest is that engineers find it very effective to write specifications in the pattern language and check their precise semantics observing the automaton visualizations. The pattern language itself is also new, and interesting in the sense that engineers are able to learn it fairly quickly, and found that it was useful for expressing most realistic properties.

The decision to analyze logs probably turned out to be crucial to the adoption of formal specifications by a testing team at JPL. Logs were already produced by the system under test, and hence no instrumentation effort on behalf of the flight software engineers was required (and, therefore, no problem with breaking expected real-time performance). This meant that the testing team could work more or less in isolation, without slowing the flight software engineers down. A great amount of research in runtime verification involves instrumentation issues. Automated code instrumentation is indeed important and interesting. However, it might not be a simple matter for software engineers to decide what to monitor and how to specify it. For example, specifying pointcuts in aspect oriented programming [24] can be a challenge. On the other hand, inserting print statements in code is the world’s most common method for debugging and understanding programs [31]. Our recommendation is, therefore, to apply formal runtime verification to logs when automated code instrumentation becomes impractical. Even unstructured logs can be made useful, though structured logging, with events as records, is generally preferable. We experimented with extracting events from text logs, using regular expressions to define events, e.g., in this example where command is bound to the number matched by (\d+):

```
event success(command) =
  'COMMAND ((\d+)) \{[^\}\]*\} SUCCEEDED'
```

We also hope to establish a connection between requirements engineering and logs, such that requirements become testable through runtime verification of logs. The common element is the event: requirements should be expressed as predicates on sequences of events, and logging should produce those events. The reactions of systems engineers we interacted with during MSL testing suggest that event-sequences may be a very natural way for high-level engineers (not limited to software engineers) working on requirements to formulate properties. Further future work includes merging the LOGSCOPE system with the RULER system [8] to obtain a unified system. A final goal is to give LOGSCOPE more exposure inside JPL and exploit its potential in other flight missions (and in MSL development as well as testing).

Thanks to members of the MSL team: Chris Delp, Dave Hecox, Gerard Holzmann, Rajeev Joshi, Cin-Young Lee, Alex Moncada, Cindy Oda, Glenn Reeves, Lisa Tatge, Hui Ying Wen, Jesse Wright, and Hyejung Yun.

---

3 Britain’s The Royal Academy of Engineering provided an award to facilitate this collaboration.
A.1. LOGSCOPE GRAMMAR

A.1.1. Lexical Elements

\( \langle \text{CODE} \rangle \rightarrow \{ : \ldots \text{Python code} \ldots : \} \)

\( \langle \text{NAME} \rangle \rightarrow \{ [a-zA-Z\_\-\_\-] [a-zA-Z\-0-9\_.\_] \}^* \)

\( \langle \text{NUMBER} \rangle \rightarrow \{ 0-9 \}^+ \)

\( \langle \text{STRING} \rangle \rightarrow \" \ldots \" \)

\( \langle \text{COMMENT} \rangle \rightarrow \langle \text{COMMENT}_1 \rangle \mid \langle \text{COMMENT}_2 \rangle \)

\( \langle \text{COMMENT}_1 \rangle \rightarrow / \ldots / \ldots / \)

\( \langle \text{COMMENT}_2 \rangle \rightarrow \# \ldots \\n \)

A.1.2. Grammar

A.1.2.1. Specifications

\( \langle \text{specification} \rangle \rightarrow \langle \text{CODE} \rangle \langle \text{monitor} \rangle^+ \)

\( \langle \text{monitor} \rangle \rightarrow \langle \text{ignore} \rangle \langle \text{monitorspec} \rangle \)

\( \langle \text{monitorspec} \rangle \rightarrow \langle \text{pattern} \rangle \mid \langle \text{automaton} \rangle \)

A.1.2.2. Patterns

\( \langle \text{pattern} \rangle \rightarrow \langle \text{name} \rangle ^* \langle \text{event} \rangle ^* \langle \text{consequence} \rangle ^* \)

\( \langle \text{event} \rangle \rightarrow \langle \text{monitor} \rangle ^* \langle \text{ignore} \rangle \langle \text{monitorspec} \rangle \)

\( \langle \text{consequence} \rangle \rightarrow \langle \text{monitor} \rangle ^* \langle \text{ignore} \rangle \langle \text{monitorspec} \rangle \)

\( \langle \text{constraint} \rangle \rightarrow \langle \text{NAME} \rangle ^* \langle \text{range} \rangle \)

\( \langle \text{range} \rangle \rightarrow \langle \text{NUMBER} \rangle \langle \text{STRING} \rangle \)

\( \langle \text{indexes} \rangle \rightarrow \langle \text{index} \rangle ^* \langle \text{index} \rangle ^* \)

A.1.2.3. Automata

\( \langle \text{automaton} \rangle \rightarrow \langle \text{automaton} \rangle ^* \langle \text{state} \rangle ^* \langle \text{initial} \rangle ^* \langle \text{hot} \rangle ^* \langle \text{success} \rangle ^* \)

\( \langle \text{state} \rangle \rightarrow \langle \text{modifier} \rangle ^* \langle \text{statekind} \rangle ^* \langle \text{NAME} \rangle ^* \langle \text{formals} \rangle ^* \langle \text{rule} \rangle ^* \langle \text{formals} \rangle ^* \langle \text{modifier} \rangle ^* \langle \text{CODE} \rangle \)

\( \langle \text{CODE} \rangle \rightarrow \langle \text{comment} \rangle ^* \langle \text{CODE} \rangle \)

\( \langle \text{comment} \rangle \rightarrow \langle \text{NAME} \rangle ^* \langle \text{arguments} \rangle ^* \)

A.1.2.4. Events

\( \langle \text{event} \rangle \rightarrow \langle \text{type} \rangle ^* \langle \text{constraint} \rangle ^* \langle \text{argument} \rangle ^* \langle \text{arguments} \rangle ^* \)

\( \langle \text{constraint} \rangle \rightarrow \langle \text{NAME} \rangle ^* \langle \text{range} \rangle \)

\( \langle \text{range} \rangle \rightarrow \langle \text{NUMBER} \rangle \langle \text{STRING} \rangle \)

\( \langle \text{symbol} \rangle \rightarrow \langle \text{CODE} \rangle ^* \langle \text{symbol} \rangle ^* \langle \text{CODE} \rangle \)

\( \langle \text{symbol} \rangle \rightarrow \langle \text{NAME} \rangle ^* \langle \text{arguments} \rangle ^* \)

\( \langle \text{arguments} \rangle \rightarrow \langle \text{NUMBER} \rangle ^* \langle \text{STRING} \rangle ^* \langle \text{NAME} \rangle ^* \)

\( \langle \text{NUMBER} \rangle \rightarrow \{ 0-9 \}^+ \)

\( \langle \text{STRING} \rangle \rightarrow \{ \}^+ \)

\( \langle \text{NAME} \rangle \rightarrow \{ [a-zA-Z\_\-\_\-] [a-zA-Z\-0-9\_.\_] \}^* \)

\( \langle \text{predicate} \rangle \rightarrow \langle \text{CODE} \rangle ^* \langle \text{predicate} \rangle ^* \langle \text{CODE} \rangle \)

\( \langle \text{CODE} \rangle \rightarrow \langle \text{comment} \rangle ^* \langle \text{CODE} \rangle \)

\( \langle \text{comment} \rangle \rightarrow \langle \text{NAME} \rangle ^* \langle \text{arguments} \rangle ^* \)
REFERENCES


