Verification of an Aerial Video System

Youcheng Sun
Dept. of Computer Science
University of Oxford, UK

Étienne André
LIPN, CNRS, UMR 7030
Université Paris 13, France

Giuseppe Lipari
CRIStAL - UMR 9189
University of Lille, France

Abstract—In this paper we propose solutions to the consolidated version of the WATERS 2015 industrial challenge of a distributed video processing system using the formalism of Parametric Timed Automata (PTA). The first challenge is harder because of the very large number of states to be analyzed, so we only provide upper bounds. The second challenge consists of a real-time scheduling problem for which we provide exact solutions by using a scheduling analysis based on the critical instant, and a PTA model.

Furthermore, regarding the second challenge, we provide formal analysis for the frequency of possible temporal violations in the system, by applying the latest results from the so-called weakly hard real-time schedulability analysis. This improvement contributes the major extension of this paper with respect to our solutions to the previous version of the challenge.

I. INTRODUCTION

With the increasing size and complexity of concurrent systems, the need for formal verification techniques has become higher and higher in the past decades. Systems mixing time and concurrency are especially subject to undesired behaviors (deadlocks, race conditions, etc.); often, their complexity makes the use of formal methods very challenging.

In this paper, we address the “consolidated WATERS industrial challenge 2015” proposed by Thales (http://waters2017.inria.fr/challenge/#Challenge17) by extending our solutions [ALS15] (to its previous version) that use the formalism of parametric timed automata (PTA). Timed automata (TA) [AD94] are a well-known formalism for specifying and verifying concurrent real-time systems. TA extend finite-state automata with a set of clocks (real-time variables growing linearly) that can be compared with integer constants. TA are used in several powerful tools such as UPPAAL [LPY97] or PAT [SLDP09]. However, the binary answer (“yes” or “no”) output by model checking is not always satisfactory; indeed, it does not allow to change or optimize some values of the system constants, nor (in general) to evaluate the system robustness, i.e., the infinitesimal variation of timing constants while preserving the reachability. PTA [AHV93] extend TA with rational-valued parameters allowed in place of constants. PTA are particularly well suited to verify systems where some timing delays are known with uncertainty. The natural drawback of PTA is the infamous state space explosion, that may prevent the verification to be truly scalable. We will use the tool IMITATOR [AFKS12] that takes PTA as an input formalism.

This work is partially supported by the ANR research program PACS (ANR-14-CE28-0002).

Improvements w.r.t. solutions in [ALS15]:

• The improvement comes from the solution to Challenge 2, where we provide also an analysis for the frequency of possible temporal violations in the video system. Although this is not explicitly specified by the challenge itself, we believe the result by natural fits into the interest of the challenge, and it offers a formal approach to evaluate the system performance.

• This extension is based on the so called weakly hard schedulability analysis, which estimates the number of deadline misses within an arbitrary time window.

Outline: We briefly review PTA and IMITATOR (Section II). Then, we recall the challenge (Section III). We derive solutions to Challenge 1A formally by using IMITATOR (Section IV). We provide a different PTA model for the much harder Challenge 1B (Section V) for which we derive an upper bound. Then, we analyze the problem in Challenge 2 (Section VI) and the improvement in the new solution is at Section VI-C. Finally, we conclude in the work (Section VII).

II. PARAMETRIC TIMED AUTOMATA

Timed automata are finite-state automata augmented with clocks, i.e., real-valued variables increasing uniformly, that are compared within guards and invariants with timing delays [AD94]. Parametric timed automata (PTA) [AHV93] extend timed automata with parameters, i.e., unknown constants, that can be used in guards and invariants.

Given a set $X$ of clocks (real-valued variables) and a set $P$ of parameters (unknown rational-valued constants), a constraint $C$ over $X$ and $P$ is a conjunction of linear inequalities on $X$ and $P$.

1. Given a parameter valuation (or point) $v$, we write $v \models C$ when the constraint where all parameters within $C$ have been replaced by their value as in $v$ is satisfied by a non-empty set of clock valuations.

Definition I: A parametric timed automaton (PTA) $A$ is $(\Sigma, L, l_0, X, P, I, E)$ with $\Sigma$ a finite set of actions, $L$ a finite set of locations, $l_0 \in L$ the initial location, $X$ a set of clocks, $P$ a set of parameters, $I$ the invariant assigning to every $l \in L$ a constraint over $X$ and $P$, and $E$ a step relation consisting of elements $(l, g, a, R, l')$, where $l, l' \in L$, $a \in \Sigma$, $R \subseteq X$ is the set of clocks to be reset, and the guard $g$ is a constraint over $X$ and $P$.

1Note that this is a more general form than the strict original definition of PTA [AHV93]; since most problems for PTA are undecidable anyway, this has no practical incidence, and increases the expressiveness of the formalism.
The semantics of a PTA $A$ can be found in, e.g., [AHV93], [AS13].

Most problems related to PTA (e.g., the parametric reachability of a location) are undecidable [AHV93], [JLR15], with some decidable syntactic subclasses related to the use of the parameters [HRSV02], [BL09], [JLR15] or on the number of clocks [AHV93], [BO14], [BBLS15] (see [And16] for a survey). We do not consider it as drawback, as we use semi-algorithms that “often” terminate in practice; we will see this is also the case for solving challenge 1A.

IMITATOR [AFKS12] is a tool for modeling and verifying systems modeled using parametric timed automata. In its latest version (2.9), IMITATOR implements several algorithms, among which:

- parametric reachability analysis: “find all parameter valuations such that some state is reachable”
- parametric robustness analysis: “given a parameter valuation $v$, find other valuations for which the discrete (untimed) behavior is the same as $v$”
- behavioral cartography: “find all parametric subspaces in which the discrete behavior is uniform”.

IMITATOR extends PTA with some useful constructions, such as constants, global discrete integer variables, strong broadcast synchronization between components, and stop-watches (i.e., the power of stopping some clocks in some locations).

The input syntax of IMITATOR is a text file, which makes it possible to write models either manually or using scripts (e.g., for large models derived from another formalism). For example, in the past several parametric schedulability analysis models were generated automatically using scripts.

### III. A Brief Description of the System

We only briefly recall the systems and the challenges; the full description is available at the workshop Web page.

#### A. Challenge 1

There are four tasks T1, T2, T3 and T4, distributed in different processing units and performing respective functionalities. The task T1 periodically receives frames from the camera and pre-processes them. Task T2 embeds further tracking information into the video frame pre-processes by Task T1. Task T2 then inserts the video frame into a register, denoted as Register23. Then Task T3 reads the frame from the register, removes the noise and tries to put the resulting video frame into a buffer, denoted as Buffer34. In the end, Task T4 reads frames from the buffer, converts them from digital to analogue and sends the final frame to the display.

Tasks T1, T3 and T4 are periodic, but their triggering clocks are subject to drift. That is, their periods $P_1$, $P_3$ and $P_4$ are unknown constants. More specifically, $P_1 \in [40 - 40 \times 0.01\%, 40 + 40 \times 0.01\%]$ms, $P_3 \in [\frac{40}{3} - 40 \times 0.05\%, \frac{40}{3} + 40 \times 0.05\%]$, and $P_4 \in [40 - 40 \times 0.01\%, 40 + 40 \times 0.01\%]$ms. Task T2 is triggered by the completion of T1.

Each task has its Best-Case and Worst-Case Execution Time (BCET and WCET) or Latency (BCL and WCL): $BCET_1 = 28\text{ms}$, $BCL_2 = 17\text{ms}$, $WCL_2 = 19\text{ms}$, $BCET_3 = WCET_3 = 8\text{ms}$. As for task T4, when it reads Buffer34 and there is no frame within the buffer, it performs an empty cycle with execution 1ms; otherwise, it executes 10ms and sends the result to display.

For such a video frame processing subsystem, we aim to tackle the following challenges.

#### Challenge 1A:

Compute the minimum and maximum latencies for a given frame from the camera output to the display input, for a buffer size $n = 1$ (challenge 1A.1) and $n = 3$ (challenge 1A.3).

Given the difficulty of Challenge 1A, we will focus on deriving the lower bound and upper bound of the end-to-end latency.

#### Challenge 1B:

Due to the different clock drifts, all frames with a same index may be discarded at the entrance of the buffer at the input of the task T4. Compute the minimum time distance between two frames produced by the camera that will not reach the display, for a buffer size $n = 1$ (challenge 1B.1) and $n = 3$ (challenge 1B.3).

#### B. Challenge 2

The complete system also includes a camera tracking subsystem that identifies objects on the camera images and commands the camera motors so to follow the objects. This subsystem consists of 3 additional tasks: Task T6 is a periodic task with period $P_6 = 100$ and it can start with a certain jitter $J_6$. Task T5 is activated by Task T6 with a synchronous call. Task T7 is activates asynchronously by Task T6 and controls the motors. Task T6 execution time is: $C_6,1 = 4\text{ms}$ before invoking Task T5; $C_6,2 \in [9, 10]$ ms after the completion of Task T5 and before the invocation of Task T7; $C_6,3 \in [4, 5]$ ms after the invocation to Task T7. Task T5 has an execution time of $C_5 \in [4, 7]$ ms. Task T7 has an execution time of $C_7 \in [11, 14]$ ms. All tasks execute on the same processor together with Task T2 (described in the first challenge), and they are scheduled by a fixed priority scheduler. Task T2 has a computation time of $C_2 = 17\text{ms}$. Finally, task priorities are assigned so that $T2 > T6 > T5 > T7$.

#### Challenge 2A:

1) Compute the best and worst-case end-to-end latencies from the activation of Task T6 to the completion of Task T7 when $J_6 = 0$. 2) Compute the best and worst-case end-to-end latencies when $J_6 = 20\text{ms}$.

#### Challenge 2B:

Tasks T2 and T5 have access to a shared mutually exclusive resource protected by the priority ceiling protocol. The access to the shared resource takes 2ms for both tasks. 1) Compute the best-case and worst-case end-to-end latencies from activation of T6 to termination of T7 for a jitter value $J_6 = 0\text{ms}$ 2) compute the best-case and worst-case end-to-end latencies from activation of T6 to termination of T7 for a jitter value $J_6 = 20\text{ms}$. 3) The optimum priority assignment minimizing the worst-case latency for a jitter value $J_6 = 0\text{ms}$ and $J_6 = 20\text{ms}$.

### IV. Solving Challenge 1A Using IMITATOR

In this part, we formally solve the challenge using PTA. At first, we will solve the case with $n = 1$ for Buffer34. The
corresponding PTA model is shown in Figure Fig. 1. Later, we show how to adapt this model to a larger buffer.

Experimental environment: We used the latest version of IMITATOR (v2.7-beta2, build 1073) with no specific modification of the tool. We just used a small Python script to parse the long list of intervals that IMITATOR outputs, and to produce a single minimum and maximum. The latest version (2.9) of the tool automatically does this, and the Python script is therefore not needed anymore. Sources, binary and models are available at [XP].

A. Camera, Task T1, Task T2

In order to reduce the state space, we model the camera, Task T1 and Task T2 into a single PTA (Fig. 1a). We also use this PTA to non-deterministically initialize the buffer and the frame currently processed by Task T4.

We choose an arbitrary frame with index target for end-to-end latency estimation and we start from the exact point such that the target frame is handled from Task T1 to task T2. A clock ckT1T2 is initialised to be WCET1 and measures the end-to-end latency of target frame. Discrete variables frame_in_3 and frame_in_4 represent the index of frames in Task 3 and Task 4 respectively. The value 0 is used to denote that there is no frame in a task. reg23 and buffer34 are for frames within Register23 and Buffer34. While we assume frame index in the system is monotonically increased, highest34 denotes the highest frame index among frames having been stored inside the buffer. For the register, buffer and Task 3, they may or may not initially contain a frame.

We do not model the period of the camera (or task 1), since we are only interested in a single frame. Let us now model the buffer. IMITATOR does not support other kinds of global variables than discrete integer variables. For a buffer with one slot (n = 1), its status can be modeled by using two discrete variables buffer34 and highest34.
• \text{buffer}_{3,4} \text{ denotes the index of current frame inside Buffer34.}

• \text{highest}_{3,4} \text{ is the highest index recorded so far.}

B. Task T3

The period of Task T3 is a \text{parameter} \text{P3\_uncertain}, that is initialized as follows:

\[
P3\_\text{uncertain} \in [40 - P3\_\text{delta}, 40 + P3\_\text{delta}]
\]

where \text{P3\_delta} = 0.05 \% \times 40 = \frac{1}{170}. Recall that parameters in PTA are unknown \textit{constants}, \textit{i.e.}, the value of which cannot evolve during the execution; this is exactly what we need to model \text{P3\_uncertain}.

Task T3 is modeled by a periodic PTA in Fig. 1b. cktT3 is a clock variable for recording task 3’s activation and execution. At the initial point, the PTA T3 is non-deterministically waiting for a new activation or executing. When T3 finishes execution, it writes into Buffer34 if the buffer is empty and its current frame has not been put into the buffer. We define a function call for this writing operation \text{write\_by\_T3}():

\[
\text{buffer}_{3,4} := \text{highest}_{3,4} := \text{frame\_in\_3}
\]

\text{IMITATOR} \text{ does not support function call in a model; here we still utilise the notion of function call for simplicity. Otherwise, task 3’s writing fails. As shown in PTA T3, we utilise the operator ”\text{ }||\text{ “ (”or”) in the edge representing writing failure. Again, this is for saving some space.}

C. Task T4

We use here a modeling mechanism similar to Task T3. The period of Task T4 is a \text{parameter} \text{P4\_uncertain}, that is initialized as follows:

\[
P4\_\text{uncertain} \in [40 - P4\_\text{delta}, 40 + P4\_\text{delta}]
\]

where \text{P4\_delta} = 0.01 \% \times P4 = 0.004.

Task T4 is modeled by a periodic PTA as in Fig. 1c. A clock variable cktT4 is used for task 4’s periodic activation and execution. When T4 activates, If the buffer is empty, T4 goes directly back to another waiting cycle. According to the system specification, there should be empty processing session for task 4. However, such an empty processing does not affect the end-to-end latency of any frame, and we omit it in PTA T4. If the buffer is not empty, task 4 reads a frame from the buffer by the function call \text{read\_by\_T4}():

\[
\text{frame\_in\_4} := \text{buffer}_{3,4}, \text{buffer}\_3,4 := 0
\]

Task T4 takes 10ms to process a frame, after the processing, if its current frame is the target one, cktT4 is reset and T4 moves to an ending location. E2E \geq 0 is the parameter for representing all possible end-to-end latencies of the target frame.

D. Deriving the Latency for \(n = 1\)

As we have seen, in order to avoid exploring the exact configurations in the system, we target a single frame (which explains the non-cyclic behavior of the PTA modeling the camera, tasks 1 and 2) that is output from the task 1 at \(t = WCET_1\). The main idea is that, at \(t = WCET_1\), the initial state must be \textit{arbitrary}, \textit{i.e.}, encode all possible configurations that could happen in the system. However, such a model may be pessimistic for containing behaviors that cannot really happen in the system. Again, we aim to derive upper and lower bounds on end-to-end latency of an arbitrary frame.

After developing the model, we use \text{IMITATOR} \text{ to perform parametric reachability analysis of location T4end\_ok, that is we ask \text{IMITATOR} \text{ to return all parameter valuations such that T4end\_ok is reachable. Then, \text{IMITATOR} hides (using existential quantification) all parameters except E2E, and then returns a list of intervals for E2E. After some post-processing to unify intervals, we get}

\[
E2E \in [63, 145.008].
\]

E. Deriving the Latency for \(n = 3\)

For the case of \(n = 3\) for \text{Buffer34}, we can keep the same \text{IMITATOR} \text{ model, with the exception of the buffer modeling. We assume that access to elements in the buffer follows a FIFO manner.}

Let us model \text{Buffer34} \text{ for } n = 3; \text{ besides the variable } \text{highest}_{3,4}, \text{ we need 3 more discrete variables for frames within each slot in the buffer: } \text{buffer}_{3,4}^1, \text{buffer}_{3,4}^2, \text{ and } \text{buffer}_{3,4}^3. \text{ When non-deterministically initializing the buffer, we need to take into account all possible scenarios of frame occupation within it. When T3 writes into the buffer, it needs to find the first free position (say } x\text{th}); \text{ thus the writing call becomes } \text{write\_by\_T3}():

\[
\text{buffer}_{3,4}^x := \text{highest}_{3,4} := \text{frame\_in\_3}
\]

Similarly, for T4 to read from the buffer, we call the adapted function \text{read\_by\_T4}():

\[
\text{frame\_in\_4} := \text{buffer}_{3,4}^1
\]

\[
\forall x \in \{1, 2\} \quad \text{buffer}_{3,4}^x := \text{buffer}_{3,4}^{x+1}
\]

\[
\text{buffer}_{3,4}^3 := 0
\]

Note that the buffer status is also updated by the reading operation such that the first slot always contains the oldest frame.

Then, we can apply parametric analysis on the model for \(n = 3\) using \text{IMITATOR}. A projection of all possible values to parameter E2E gives the following result:

\[
E2E \in [63, 225.016].
\]

Finally, we conclude results for Challenge 1A in Table I.
TABLE I: E2E latency results for Challenge 1A

<table>
<thead>
<tr>
<th>Buffer34 size</th>
<th>min E2E</th>
<th>max E2E</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 1</td>
<td>63 ms</td>
<td>145.008 ms</td>
</tr>
<tr>
<td>n = 3</td>
<td>63 ms</td>
<td>275.016 ms</td>
</tr>
</tbody>
</table>

V. SOLUTION TO CHALLENGE 1B

Due to the page limit, please refer to our solutions [ALS15] to the WATERS 2015 industrial challenge for more details.

VI. SOLUTION TO CHALLENGE 2

A. Schedulability analysis

For Challenge 2A, when it goes to maximum end-to-end latency, we could employ the critical instant property [LL73], [LSAF14] to compute it. That is, the maximum latency happens such that

- T5, T6 and T7 always execute by their worst-case.
- T6 starts to execute coincidently with a release of T2.

We assume the period of T2 is exactly equal to $P_2 = 40$ ms; Then, we obtain the maximum latency 74 ms for $J_6 = 0$ and $74 + 20 = 94$ ms for $J_6 = 20$ ms. Notice that they are both inferior to $P_6$, so there is no possibility that a new instance of T6 starts before the last instance of T7 completes. Possible variations in the period of T2 do not have any impact on the estimated end-to-end latency.

As for the minimum latency, a first intuition is that T5, T6 and T7 should execute by their best-case, so we compute $4 + 4 + 9 + 4 + 11 = 32$ ms.

- If the initial offset between T2 and T6 is larger than 32 ms, then the minimum latency is indeed 32 ms.
- Otherwise, T2 preempts the execution of T5-T7 at least once, and the minimum latency would be $32 + 17 = 49$ ms.

Therefore, we conclude that the latency of the chain T5-T7 is within [49, 94].

In Challenge 2B, a mutually exclusive resource is shared between T5 and T2. In this case the minimum and maximum latencies do not change, whereas task T2 could suffer for a delay of 2 in accessing the shared resource, so its response time will vary in [17, 19].

If we have the freedom to manipulate priorities we can further reduce the end-to-end latency. T2 is part of the chain for video frame processing, and to avoid interfering the verification result in Challenge 1, we still assume that T2 has higher priority than T5, T6 and T7. We can re-arrange the priorities among T5, T6 and T7 as $T_5 > T_7 > T_6$, and the worst-case latency becomes 69 ms for $J_6 = 0$ and 89 ms for $J_6 = 20$ ms. It is not difficult to see that this is the optimum priority assignment by enumeration.

B. PTA model for Challenge 2

We modeled the system following the scheduling framework for PTA proposed in [LSAF14]. We do not report here the full model for lack of space, it can be found at [XP]. We just summarize the main points.

- The scheduler is modeled as a separate automaton, generated automatically from the priority ordering, and it interacts with the task automata using synchronization.
- Initial task offsets are modeled as non-deterministic choices: the period of T6 is $p_6 \in [0, P_6 - 32]$, whereas the period of T2 is $p_2 \in [0, P_2 - 17]$.
- An observer automaton measures the end-to-end latency with a clock. Maximum and minimum latencies are modeled as parameters, exactly in the same way as the model developed for Challenge 1A.

In the end, we report the results for Challenge 2 in Table II. As we discussed, the existence of a shared resource between T5 and T6 does not affect the latency for the chain T5-T7.

<table>
<thead>
<tr>
<th>min latency</th>
<th>max latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>J6 = 0 ms</td>
<td>49 ms</td>
</tr>
<tr>
<td>J6 = 20 ms</td>
<td>94 ms</td>
</tr>
</tbody>
</table>

(a) The min/max latency

TABLE II: Results for Challenge 2

C. The weakly hard analysis for Challenge 2

This part contains the improvement in our new solutions with respect to [ALS15]. Regarding the optimal priority assignment, the scheme in Section VI-A has to maintain that task T2 is assigned the highest priority, otherwise, T2 cannot be guaranteed to complete the execution before its next activation, which indicates the violation of the (implicit) temporal constraint. Different from the previous solution, this time we respect the optimal priority assignment scheme that minimizes the end-to-end latency for Challenge 2: $T_5 > T_7 > T_6 > T_2$. Instead, we are going to formally analyze how often T2 violates its temporal constraint, and this is the so called weakly hard analysis.

The weakly hard analysis of a task is also named $m$-$K$ analysis, which in principle aims to check if there will be more than $m$ temporal violations out of arbitrarily $K$ consecutive activations of a task.

Three kinds of the weakly hard analysis exist. [BBL01] is the first work that studies the $m$-$K$ model, however, it relies on the explicit system initial status (i.e., the first release time of every task is predefined) and the periodicity of a periodic task system for the weakly hard analysis. More recent works based on the typical worst-case scenario [QHE12] break the limitation on predefined initial task release times, and they are eligible for weakly hard analysis of systems in the case of the occasional sporadic workload.

Both the two approaches above do not fit into Challenge 2, as it is a periodic system with unknown task release times and there is no sporadic workload. Instead, we apply the tool
Weakly-hard Schedulability Analyzer (WSA) [WSA], which implements latest results for the $m$-$K$ analysis. Different from other methods,

- WSA is able to conduct weakly hard analysis on general periodic systems, even when initial task release times are unknown;
- it supports tasks with release jitters;
- it supports the resource sharing, e.g., the use of Priority Ceiling Protocol (PCP) [SRL90];
- its analysis is based on Mixed Integer Linear Programming, and is compatible with parametric task configurations (like the period of task T2).

Finally, for the weakly hard analysis of task T2, we abstract T5, T6 and T7 into a single task T6' that has period 100ms, best-case execution time 32ms, worst-case execution time 40ms and a jitter value $j$ (0ms or 20ms). T6' has a higher priority than T2. By applying WSA to T2, the weakly hard analysis results obtained are in Table III, where we list the maximum number ($m^*$) of temporal violations of T2, within an arbitrary sequence of $K$ its activations. In this particular case, the jitter value does not change the weakly analysis result.

\[
\begin{array}{|c|c|}
\hline
m^* & K \\
\hline
1 & 2 \\
2 & 5 \\
4 & 10 \\
6 & 15 \\
\hline
\end{array}
\]

**TABLE III**: The weakly hard analysis results of Challenge 2

According to the results in Table III, there will be at most 2 temporal violations for T2 every 5 its activations and this figure is consistent with the result when $K = 10$ and 15. Also note that, in Table III, when $K = 2$, $m^* = 1$. This implies that there will be never two consecutive temporal violations that happen in a row.

VII. CONCLUSION

We proposed solutions, using formal methods based on PTA, to the consolidated version of the WATERS 2015 industrial challenge. Thanks to their expressiveness, PTA are especially convenient for modeling systems that contain some unknown but constant configurations. Challenges 1A and 1B were the most difficult, so we only provided upper bounds on results for them. The high complexity is mainly due to the large hyperperiods, therefore a complete model is intractable. We spent about one week for studying the problem and provide a model that could converge in a decent time.

IMITATOR is the software tool we rely on for performing parametric analysis. IMITATOR turned out to be particularly well-suited for the analysis of such systems with period constants but with some (unknown) imprecision.

Conversely, Challenge 2 is a scheduling problem, whose solution can be obtained quickly using schedulability analysis methods. The solution has been validated by a formal model based on PTA that took half a day to be built using the framework developed in [LSAF14], and only a few seconds to converge.

We further apply the weakly hard analysis tool WSA to evaluate the frequency of temporal violations that appear in Challenge 2, which helps elaborate more design choices for the system.

REFERENCES


